Bayesian Estimation

Publishing Globally

S. K. SINHA

NEW AGE INTERNATIONAL PUBLISHERS

Bayesian Estimation

S.K. SINHA
Professor of Statistics
University of Manitoba
Winnipeg, Canada

PUBLISHING FOR ONE WORLD
NEW AGE INTERNATIONAL (P) LIMITED, PUBLISHERS
New Delhi · Bangalore · Calcutta · Chennai · Guwahati · Hyderabad
Lucknow · Mumbai · Pune

NEW AGE INTERNATIONAL (P) LIMITED, PUBLISHERS

NEW DELHI	:	4835/24, Ansari Road, Daryaganj, New Delhi-110 002
BANGALORE	:	No. 36, Mallikarjuna Temple Street, Basavangudi Bangalore-560 004
CALCUTTA	:	37-G, Bondel Road, Calcutta-700 019.
CHENNAI	:	20, IInd Main Road Kasthuribai Nagar, Adyar, Chennai-600 020
GUWAHATI	:	Pan Bazar, Rani Bari, Guwahati-781 001
HYDERABAD	:	1-2-412/9, Gaganmahal, Near A.V. College, Domalguda, Hyderabad-500 029
LUCKNOW	:	18, Madan Mohan Malviya Marg, Lucknow-226 001
MUMBAI	:	128/A, Noorani Building, Block No. 3, First Floor, L.J. Road, Mahim, Mumbai-400 016.
PUNE	:	44, Prashant Housing Society, Lane No. 6, Paud Road, Kothrud, Pune-411029.

ISBN 81-224-1171-1

Published by H.S. Poplai for New Age International (P) Limited, 4835/24, Ansari Road, Daryaganj, New Delhi-110002. Typeset by Jayanti Graphics and Printed at Ram Printograph (India), Okhla Industrial Area, New Delhi-110020.

Printed in India

Production : M.I. Thomas

Dedicated to the memory of Professor Morris H. DeGroot — an eminent scientist, a good friend, philosopher and guide and, above all, a great human being.

Foreword

Professor Sinha has made significant contributions in the field of Bayesian Inference. We both share a conviction that the Bayesian approach to Statistics should now be recognised as a powerful alternative to the classical or frequentist approach.

This work, *Bayesian Estimation,* by Professor Sinha, seeks to inform us why this should be so and is a valuable contribution to Statistics, bringing together both the interpretation and application of the Bayesian approach. Detailed mathematical proofs which students at this level like to see in an introductory text such as this, are included. Judging from the success of Professor Sinha's earlier publication, *Reliability and Life Testing,* I am sure this book will be well received by students, researchers and practising statisticans. I congratulate him on its seeing the light of day.

State University of New York,
Buffalo, NY, USA

IRWIN GUTTMAN

Preface

I have been working in the area of Bayesian Inference for quite sometime and along with many of my friends and colleagues, who share my research interests, I believe there is a real need for a suitable book on Bayesian Statistics which will be appropriate for one-semester course for advanced undergraduate or first year graduate students of Statistics, Econometrics, Management and Engineering Sciences. This book is an attempt to present the theory and methodology of Bayesian Estimation and is primarily based on material covered by me in a semester long first year graduate course in Bayesian Inference at the University of Manitoba, Canada, Panjab University, Chandigarh, India, and in subsequent short courses elsewhere in this area. It is assumed that students have calculus background and have completed a course in Mathematical Statistics including standard distribution theory and introduction to the general theory of estimation.

The book has eight Chapters and an Appendix with eleven sections. Chapter 1 reviews elements of Bayesian paradigm. Chapter 2 deals with Bayesian estimation of parameters of well-known distributions, *viz.*, Normal and associated distributions, Multinomial, Binomial, Poisson, Exponential, Weibull and Rayleigh families. Chapter 3 considers predictive distributions and predictive intervals. Chapter 4 covers Bayesian interval estimation. Chapter 5 discusses Bayesian approximations of moments and their application to multiparameter distributions. Chapter 6 treats Bayesian regression analysis and covers linear regression, joint credible region for the regression parameters and bivariate normal distribution when all parameters are unknown. Chapter 7 considers the specialised topic of mixture distributions (which may be omitted in the first reading) and Chapter 8 introduces Bayesian Break-even Analysis.

The Appendix reviews the results used in text. A number of examples have been worked out at the end of each section and a set of exercises given at the end of each chapter. The text lists-excellent references ranging from 1931 to 1993 which may encourage the readers to

undertake further research in this interesting area. Researchers in various disciplines, instructors, Reliability engineers and consultant statisticians will also find this publication a helpful guide.

I owe a great deal to my students for their comments, criticism and suggestions on various topics at the preparatory stage of the manuscript, to my colleagues for long hours of discussions and debates which often ended in disagreement, to the referee for the valuable remarks and excellent feedback on this proposal and the series of seminars on Bayesian Inference in Econometrics and Statistics organised by Prof Arnold Zellners. I would like to express my gratitude to Professor D.V. Lindley, J.B. Kadane, James Press and Prem Goel for their stimulating advice and assistance at the launching of the project. I am particularly indebted to late Professor M.H. Degroot for his constant encouragement and overall interest in this monograph. The writing of this text started at the University of Manitoba and continued and was completed during my 1992 sabbatical tenures at Pennsylvania State University and Panjab University, Chandigarh, India. I gratefully acknowledge the prolonged discouses on Bayesian mode of reasoning with Professors Charles Antle of Penn-State and Harshinder Singh at Chandigarh. Their warm hospitality and research support will never be forgotten.

I am grateful to Mesers Allyn and Bacon, Richard Irwin, McGraw-Hill and Wadsworth for permission to include Examples 1.7, 8.1, 8.6-8.9 and Exercises 2, 3-7, 10-12 in Chapters 1 and 8.

It gives me great pleasure to acknowledge the partial financial support for this project provided by the University of Botswana Research and Publication Committee while I was at University of Manitoba/ University of Botswana Link Professor of Statistics during 1993-95. I am indebted to Ms Mabel Davies of University of Manitoba for her constant care, patience and skill in typing the manuscript, to my secretary Ms Duna Molaolwa for her support and co-operation while I was at the University of Botswana, to my colleague Dr. J.A. Sloan, Nursing Research Institute, University of Manitoba for computing assistance and research contribution, and to New Age International (P) Limited, publishers for their continued co-operation and interest in my work.

I do not find enough words to express my appreciation for the all-out support of my wonderful family without which this text would have never been completed.

Thanks a million Ruby, Pamela and Debashis.

S.K. SINHA

Contents

CHAPTER

1 Elements of the Bayesian Paradigm

1.1 DEFINITION AND INTERPRETATION OF PROBABILITY

Probability plays an important role in almost every field — industry, commerce, physical and biological sciences, statistics and of course, in our daily life. Whenever we make a decision under uncertainty, consciously or otherwise, we actually make a probability statement. Your investment broker calls you up and advises you to invest $ 1000.00 on a certain stock. If you invest, you may make a profit or may break even and get your money back, or may suffer a considerable loss. Should you buy a $15.00 ticket for a 2 million dollar raffle? If you do and win, you will be a very rich man. If you buy, and do not win, you will lose $ 15.00. If you do not buy, you will not lose $ 15.00, but you could have won had you bought the ticket. You have a parcel of riverside lots and the real estate market is booming. You wonder, "Should I sell the property now or wait for the market to rise further up?"

What are the respective probabilities? What is the probability that you will make money in the stock market or you will win the raffle or the real estate will further rise? How can one calculate the probability that a certain event will occur? In some problems such as 'What is the probability of obtaining a head in a single toss of a fair coin' or 'What is the

probability that a 6 will show up when a fair die is rolled', the answers 'The corresponding probabilities are 1/2 or 1/6' are almost intuitive, without any knowledge of a complicated formula for computing probabilities. Now suppose we ask, 'What is the probability that a card drawn at random from a full pack will be an ace'? Perhaps the answer will not be as instantaneous as in the other two cases.

In order to provide a reasonable answers to questions such as these, we need a definition of probability. Collins Concise Dictionary defines probability as a measure of the degree of confidence one may have in the occurrence of an event in a scale from zero (impossibility) to one (certainty). But this does not help us. All it tells us is that probability is a ratio which cannot be less than zero or greater than one. We look for a more workable definition of probability, one that will give us a numerical measure of probability that an event A will occur. We consider the following definitions of probability:

Classical definition

The classical definition, due to Pierre Simon Laplace (1749–1827) is as follows:

If an event can take place in n mutually exclusive ways, *all equity likely*, and if m of these corresponds to what we would call "success", then probability of success in a single trial is the ratio $\frac{m}{n}$.

Thus, for the card problem, each one of the 52 cards in the pack is equally likely to be drawn and there are 4 aces in the pack. We have $n = 52$, $m = 4$, and hence, the probability of an ace $= \frac{4}{52}$.

Karl Pearson, in his History of Statistics in the 17th and 18th Centuries (edited by E.S. Pearson, 1978), remarks that the assumption 'all events are equally likely', is as fallacious as arguing, 15 people in a room attending his lecture, are equally likely to die within the next 5 years. It cannot be logically applied to a problem unless certain hypotheses are satisfied. In the context of the preceding problems, we must assume that the coins are fair, the die is unbiased and the pack of cards is well-shuffled.

The frequency definition

Sampling theorists interpret probability as the relative frequency of success in a conceptually infinite sequence of independent trials, all made under *identical conditions*.

Let r be the number of observed successes in such a sequence of n independent trials and let p be the true probability of success. Then the relative frequency of success,

$$\frac{r}{n} \to p \quad \text{as} \quad n \to \infty.$$

Let U be a non-negative random variable and let $E(U) = a$. From Chebyscheff's inequality,

$$P\,[U \leq at^2\,] \geq 1 - \frac{1}{t^2}$$

it follows that,

$$P\left[\,\left|\frac{r}{n} - p\right| \leq \epsilon\,\right] \geq 1 - \frac{p\,(1-p)}{n\,\epsilon^{\,2}}.$$

$$\lim_{n\to\infty} P\left[\,\left|\frac{r}{n} - p\,\right| \leq \epsilon\,\right] = 1 \quad \text{or}$$

$$\lim_{n\to\infty} P\left[\,\left|\frac{r}{n} - p\right| > \epsilon\,\right] = 0 \qquad \text{for all } \epsilon > 0.$$

which statement implies that the relative frequency of success tends to the true probability p as the number of trials is indefinitely increased.

Kenny and Keeping (1951) point out that the frequency definition of probability also has certain deficiencies. For example, how can one guarantee that the trials (or experiments) have been repeated under identical conditions? Consider a case of buying a life insurance. The probability that a person aged 50 and who passed the prescribed medical test will die within a year is assessed from the birth and death statistics (at various age distributions) recorded by Government agencies or large Insurance companies. Clearly the conditions have not remained constant. Due to improved hygiene and advanced medical facilities, there has been steady decrease in mortality rates.

Degree of belief approach

In several instances the frequency definition of probability defies a logical interpretation and a 'subjective' or 'degree of belief' interpretation makes more sense than the frequentist approach. Subjective probabilists, better known as Bayesians, interpret probability as a measure of a person's degree of rational belief in a proposition (Keynes, 1921; Ramsey, 1931; De Finneti 1937; Savage, 1954, 1961, 1962; Jeffreys, 1961). Suppose John claims, "I am 80% certain that I will win the scholarship". What does 80% mean? In the 'degree of belief' context it means that John has the same degree of confidence in his winning the scholarship as he would in the proposition that when a ball is picked up at random from an urn containing 8 white and 2 black balls, the ball win turn out to be white. It is impractical to imagine that if the scholarship contest is repeated, say 1000 times, John will be successful in 800 contests and will be unsuccessful in the remaining 200.

Following Lindley (1965), another interpretation is that John will be prepared to offer a bet at odds 8 to 2 in which he will pay 8 units if he does not win the scholarship and receive 2 units if he does.

The 'degree of belief' in a certain proposition A depends on the state of our cumulative knowledge about A and is, therefore, almost always a conditional probability $P(A/B)$ where B represents out current information relevant to A. As our accumulated knowledge about A changes, so does our 'degree of belief' in A (Lindley, 1965).

Axiomatic approach

An event A is represented by a subset of the 'set of all possible outcomes of an experiment' called the sample space. Let S be a sample space with n (finite) elements ($E_1, E_2, \ldots, E_n$); we have $\binom{n}{1}$ number of subsets each containing one element,

$\binom{n}{2}$ number of subsets each containing two elements,

$\vdots$

$\binom{n}{n} = 1$ subset containing all the elements, which is the sample space S itself, and finally we have the subset containing no element or the empty set ϕ.

Thus, the total number of subsets in the sample

$$= 1 + \binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{n} = 2^n.$$

Each of these subsets defines an event.

Let A_1 and A_2 be two subsets in the sample space S.

The union of the sets A_1 and A_2 represented by $A_1 \cup A_2$ is the set of all points x which belong to A_1 or A_2 or both.

Let $\quad A_1 = \{ 1, 2, 8, 7 \}$

Let $\quad A_2 = \{ 1, 3, 7, 9, 5 \}$

Then, $\quad A_1 \cup A_2 = \{ 1, 2, 3, 5, 7, 8, 9 \}$.

The intersection of the sets A_1 and A_2 represented by $A_1 \cap A_2$ is the set of all points y which belong to A_1 as well as A_2.

Thus, $\quad A_1 \cap A_2 = \{ 1, 7 \}$.

Suppose to each event A_i in the sample space S we assign a number $P(A_i)$ such that

(i) $0 < P(A_i) < 1$

(ii) $P(A_i \cup A_j) = P(A_i) + P(A_j)$ if $A_i \cap A_j = \phi$,

and

(iii) $P(S) = 1$ (or $P(\phi) = 0$)

then $P(A_i)$ is called the probability of the event A_i.

Consider the experiment of rolling a pair of fair dice. The experiment generates a sample space with 36 elements (i, j),

where $\quad i =$ dice no. 1 shows the point i and

$j =$ dice no. 2 shows the point j, $\quad i, j = 1, 2, \ldots, 6$.

The elements of the sample space are:

1, 1 2, 1 3, 1 4, 1 5, 1 6, 1

1, 2	2, 2	3, 2	4, 2	5, 2	6, 2
1, 3	2, 3	3, 3	4, 3	5, 3	6, 3
1, 4	2, 4	3, 4	4, 4	5, 4	6, 4
1, 5	2, 5	3, 5	4, 5	5, 5	6, 5
1, 6	2, 6	3, 6	4, 6	5, 6	6, 6

Let us assign a number $\frac{1}{36}$ to each of these elements and let A be the event that the sum of the points shown on the faces of the dies equals 8. Then,

$$A = \{ 2,6; 3,5; 4.4; 5,3; 6.2 \}$$

$$\equiv \{ A_1;\ A_2;\ A_3;\ A_4;\ A_5 \}$$

Thus $$P(A) = \sum_{i=1}^{5} P(A_i) = \frac{5}{36}.$$

Definitions

If the events A and B are such that they cannot occur together, then they are called mutually exclusive events. In the set-theoretic context, $A \cap B = \phi$, and hence, $P(A \cap B) = 0$.

Let A' be the set of all points that do not belong to A'. Such a set A' is called the complement of A. If A is an event, A' represents the event not $-A$.

1.2 LAWS OF PROBABILITY

(1) Addition Law

Theorem: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Proof: Since the event A may occur either B or without B, we have

$$A = (A \cap B) \cup (A \cap B')$$

Since $(A \cap B)$ and $(A \cap B')$ are mutually exclusive, from (ii)

$$P(A) = P(A \cap B) + P(A \cap B').$$

Similarly,

$$P(B) = (B \cap A) + P(B \cap A').$$

$$P(A) + P(B) = P(A \cap B) + \{P(A \cap B') + P(B \cap A') + P(A \cap B)\}$$

$$= P(A \cap B) + P(\text{either } A \text{ or } B \text{ or both } A \text{ and } B \text{ occurring})$$

$$= P(A \cap B) + P(A \cup B).$$

Thus, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Note that if $(A \cap B) = \phi$

$$P(A \cup B) = P(A) + P(B) \quad \text{as in (ii)}.$$

Corollary: $P(A') = 1 - P(A)$

Since $A \cup A' = S$

$$P(A \cup A') = 1$$

i.e. $P(A) + P(A') = 1$

or $P(A') = 1 - P(A)$

(2) Multiplication Law

Theorem : $P(A \cap B) = P(A)P(B \mid A)$.

where $P(B \mid A)$ = Probability of B, given that A has already occurred.

Proof. Consider the following 2×2 Table:

Habits / Health	A	A′	Total
B	a	b	$a+b$
B′	c	d	$c+d$
To.al	$a+c$	$b+d$	n

Suppose n people were interviewed and classified according to A: Smoker, A': Non-smokers and B: healthy and B': evidence of lung cancer. Let a, b, c, d be the cell-frequencies, and

$a + b + c + d = n$. Then,

$$P(A) = \frac{a+c}{n}.$$

$$P(A \cap B) = \frac{a}{n}.$$

What does $\frac{a}{a+c}$ represent?

Clearly $\frac{a}{a+c}$ represents the probability of B given A or assuming A has already happened which we write as $P(B|A) = \frac{a}{a+c}$.

$$P(A)P(B|A) = \left(\frac{a+c}{n}\right)\left(\frac{a}{a+c}\right) = \frac{a}{n} = P(A \cap B).$$

Note that $P(B|A) = P(B)$ implies that the occurrence of B does not depend on A, i.e. A and B are independent. In that case,

$$P(A \cap B) = P(A)\ P(B).$$

Example 1.1:

A card is drawn at random from a full pack. What is the probability that it is a (i) king (ii) heart (iii) king or heart?

(i) $P(\text{king}) = \frac{4}{52}$;

(ii) $P(\text{heart}) = \frac{13}{52}$;

(iii) $P(\text{king} \cup \text{heart}) = P(\text{king}) + P(\text{heart}) - P(\text{king} \cap \text{heart})$

$$= \frac{4}{52} + \frac{13}{52} - \frac{1}{52}$$

$$= \frac{16}{52}.$$

Example 1.2:

An urn contains 8 white and 10 black balls. A ball is drawn at random and then returned. A second ball is drawn. What is the probability that the first ball was white and the second ball was black?

Since the first ball was returned to the urn, the color of the second ball did not depend on what the color of the first ball was.

Let A be the event that the first ball was white and B be the event that the second ball was black $P(A \cap B) = P(A)\,P(B)$.

$$= \frac{8}{18} \cdot \frac{10}{18} = \frac{20}{18}\,.$$

Example 1.3:

Urn I contains 10 white and 15 black balls. Urn II contains 15 white and 12 black balls. One ball was transferred at random from Urn I to Urn II, colour unnoted. Then a ball was drawn from Urn II. What is the probability that this ball was white?

The color of the ball from Urn II depends on the color of the ball transferred from Urn I to Urn II.

Let A_1: the even that the ball transferred from Urn I to Urn II was white and

A_2: the event that it was black.

Let A: the event that the ball drawn from Urn II was white. Then,

$$A = (A_1 \cap A) \dot{\cup} (A_2 \cap A)$$

$$P(A) = P(A_1 \cap A) + P(A_2 \cap A)$$

$$= P(A_1)\,P(A \mid A_1) + P(A_2)\,P(A \mid A_2)$$

$$= \frac{10}{25} \cdot \frac{16}{28} + \frac{15}{25} \cdot \frac{15}{28}$$

$$= \frac{11}{20}\,.$$

Conditional probability

From the multiplication law,

$$P(A \cap B) = P(A)\,P(B \mid A) \text{ or}$$

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)}, \; P(A) > 0\,.$$

$P(B \mid A)$ is interpreted as the conditional probability of B or the probability of B, given A or conditional on A.

Example 1.4:

A card is drawn from a well-shuffled full pack. Suppose it is given that the card drawn was a heart. What is the probability that it is the king of hearts?

Let H be the event that the card is a heart and K be the event that it is a king. We want $P(K \mid H)$.

$$P(K \mid H) = \frac{P(K \cap H)}{P(H)} = \frac{\frac{1}{52}}{\frac{13}{52}} = \frac{1}{13}.$$

Since it is known that the card drawn was a Heart, this probability may easily be computed by considering the restricted sample space made up of 13 elements, viz, the 13 hearts with one king.

Hence, $P(K \mid H) = \frac{1}{13}$

We will find many more examples of conditional probability in the next section on Bayes' Theorem.

Example 1.5:

Consider the following problems:

(i) Suppose we have three urns U_1, U_2, U_3 and the contents of the urns are:

U_1 : 3 white, 4 black, 5 red balls

U_2 : 5 white, 3 black, 12 red balls

U_3 : 6 white, 10 black, 4 red balls.

One urn is selected at random and a ball drawn. It turns out to be white. What is the probability that it came from U_1?

Let B_1 be the event that U_1 was chosen

B_2 be the event that U_2 was chosen

B_3 be the event that U_3 was chosen

and A be the event that the ball drawn was white. We want

$$P(B_1 \mid A) = \frac{P(A \cap B_1)}{P(A)} = \frac{P(B_1)P(A \mid B_1)}{P(A)}.$$

Now A may occur only with B_1, B_2 or B_3.

Hence, $$A = (A \cap B_1) \cup (A \cap B_2) \cup (A \cap B_3)$$

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + P(A \cap B_3)$$

$$= \sum_{i=1}^{3} P(B_i)P(A \mid B_i)$$

Thus,

$$P(B_1 \mid A) = \frac{P(B_1)P(A \mid B_1)}{\sum_{i=1}^{3} P(B_i)P(A \mid B_i)}$$

Note that $P(B_i)$ and $P(A \mid B_i)$, $i = 1, 2, 3$ are known.

$$P(B_1) = P(B_2) = P(B_3) = \frac{1}{3};$$

$$P(A \mid B_1) = \frac{3}{12}, P(A \mid B_2) = \frac{5}{20}, P(A \mid B_3) = \frac{6}{20}$$

We have,

$$P(B_1 \mid A) = \frac{\frac{1}{3} \cdot \frac{3}{12}}{\frac{1}{3} \cdot \frac{3}{12} + \frac{1}{3} \cdot \frac{5}{20} + \frac{1}{3} \cdot \frac{6}{20}}$$

$$= \frac{5}{16}$$

(ii) A purse contains 5 quarters and one of them is double-headed. A coin is chosen at random and tossed 5 times. Every time a head turns up. What is the probability that it is the quarter with two heads?

Let B_1 : the event that it is a double-headed quarter;

B_2 : the event that it is a regular quarter with a head on one side and a tail on the other.

A : the event that the coin was tossed 5 times and every time a head showed up.

We want $P(B_1 \mid A) = \dfrac{P(A \cap B_1)}{P(A)} = \dfrac{P(B_1)P(A \mid B_1)}{P(A)}$.

$$P(B_1) = \frac{1}{5}, P(B_2) = \frac{4}{5}, P(A \mid B_1)$$

$$= 1 \text{ and } P(A \mid B_2) = \left(\frac{1}{2}\right)^5.$$

$$P(A) = P(A \cap B_1) + P(A \cap B_2)$$

$$= P(B_1)P(A \mid B_1) + P(B_2)P(A \mid B_2).$$

$$P(B_1 \mid A) = \frac{P(B_1)P(A \mid B_1)}{P(B_1)P(A \mid B_1) + P(B_2)P(A \mid B_2)}$$

$$= \frac{\frac{1}{5} \cdot 1}{\frac{1}{5} \cdot 1 + \frac{4}{5}\left(\frac{1}{2}\right)^5}$$

$$= \frac{1}{1 + \frac{1}{8}} = \frac{8}{9}.$$

The exarnples (i) and (ii) are applications of a famous theorem due to Reverend Thomas Bayes (1763) which we state and prove in the following.

1.3 BAYES' THEOREM

Suppose an event A can occur only if one of the mutually exclusive and exhaustive events $(B_1, B_2, \ldots, B_k)$ occurs. Suppose the probabilities $P(B_i)$ and $P(A|B_i)$, $i = 1, 2 \ldots, k$ are known a-priori. How does the information that A has already occurred influence the probability of B_i?

We are interested in the conditional probability $P(fB_i|A)$.

This probability is given by

$$P(B_i \mid A) = \frac{P(B_i)P(A \mid B_i)}{\sum_{i=1}^{k} P(B_i)\,P(A \mid B_i)}$$

Proof. $P(A \cap B_i) = P(A)\,P(B_i \mid A)$

$$P(B_i \cap A) = P(B_i)\,P(A \mid B_i)$$

$$P(A \cap B_i) = P(B_i \cap A).$$

Hence, $P(A)\,P(B_i \mid A) = P(B_i)\,P(A \mid B_i)$ and

$$P(B_i \mid A) = \frac{P(B_i)\,P(A \mid B_i)}{P(A)} \tag{1.1}$$

Since A can occur only if B_1 or B_2 or .... B_k occurs, it follows that

$$A = (A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_k).$$

Since $B_1, B_2, \ldots, B_k$ are mutually exclusive and exhaustive,

$$P(A) = \sum_{i=1}^{k} P(A \cap B_i) = \sum_{i=1}^{k} P(B_i)\,P(A \mid B_i)$$

Substituting in (1.1), we have

$$P(B_i \mid A) = \frac{P(B_i)\,P(A \mid B_i)}{\sum_{i=1}^{k} P(B_i)\,P(A \mid B_i)}$$

Example 1.6:

A coin was lost from a purse full of change containing 10 quarters, 15 dimes and 5 nickels. 2 coins were drawn from the remaining contents in the purse and they both turned out to be quarters. What is the probability that the missing coin was a quarter?

Let B_1 : the event that the missing coin was a quarter?

Let B_2 : the event that the missing coin was other than a quarter

A : the event that the 2 coins drawn were quarters,

We want $P(B_1 \mid A)$.

$$P(B_1) = \frac{10}{30} = \frac{1}{3}$$

$$P(B_2) = 1 - \frac{1}{3} = \frac{2}{3}.$$

$$P(A \mid B_1) = \frac{\binom{9}{2}}{\binom{29}{2}}$$

$$P(A \mid B_2) = \frac{\binom{10}{2}}{\binom{29}{2}}$$

$$P(B_1 \mid A) = \frac{P(B_1)P(A \mid B_1)}{P(B_1)P(A \mid B_1) + P(B_2)P(A \mid B_2)} = \frac{2}{7}.$$

Suppose we now ask, "What is the probability that a second event C will occur, assuming that, (i) the event A has already occurred and (ii) the conditions on the events A and $\{B_i\}$, $i = 1, 2, \bullet\bullet\bullet\, k$ hold?"

We are interested in the conditional probability $P(C \mid A)$ and since A occurs only if one of the mutually exclusively and exhaustive events $\{B_i\}$ occurs, it follows that C will occur only with one of these B_i's. Hence,

$$C = (B_1 \cap C) \cap (B_2 \cap C) \cup \cdots \cup (B_k \cap C)$$

$$(C/A) = (B_1 \cap C/A) \cup (B_2 \cap C/A) \cup \cdots \cup (B_k \cap C/A)$$

$$P(C/A) = \sum_{i=1}^{k} P(B_i \cap C/A).$$

$$= \frac{\sum_{i=1}^{k} P(B_i \cap C \cap A)}{P(A)}.$$

$$P(B_i \cap C \cap A) = P(B_i \cap A \cap C)$$

$$= P(B_1 \cap A)P(C/B_i \cap A)$$

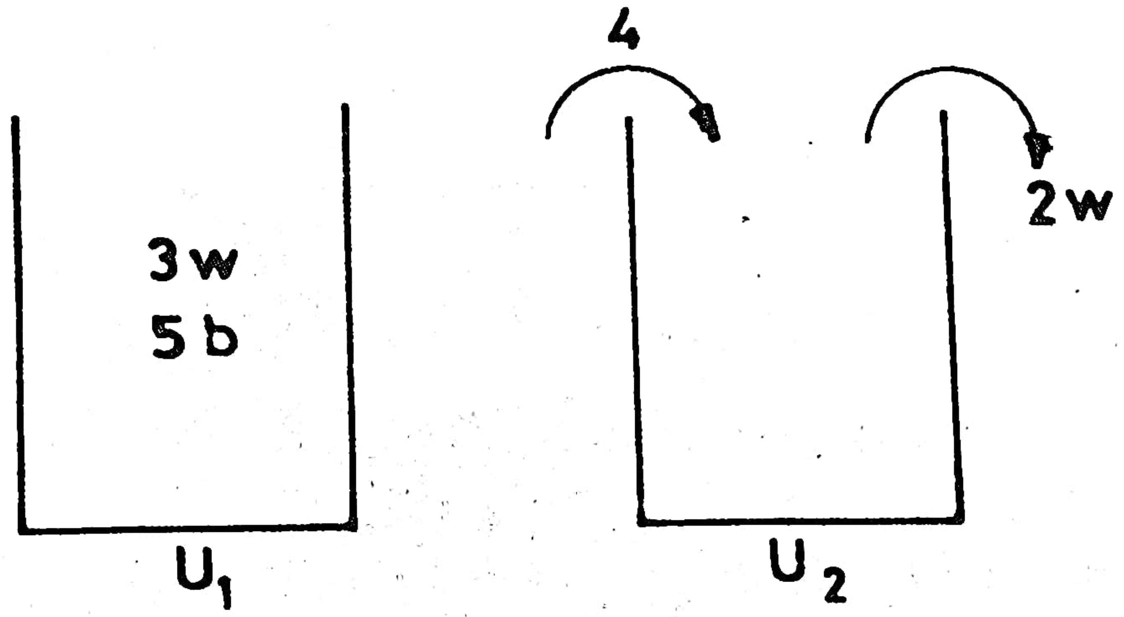
$$= P(B_i)P(A/B_i)P(C/B_i \cap A).$$

Hence, $$P(C/A) = \frac{\sum_{i=1}^{k} P(B_i)P(A/B_i)P(C/B_i \cap A)}{\sum_{i=1}^{k} P(B_i)P(A/B_i)}$$

It may happen that $P(C/B_i \cap A) = P(C/B_i)$ which implies that the event B_i has occurred and it may make C independent of A. In that case,

$$P(C/A) = \frac{\sum_{i=1}^{k} P(B_i)P(A/B_i)P(C/B_i)}{\sum_{i=1}^{k} P(B_i)P(A/B_i)}$$

*Example 1.7:



Suppose an urn U_1 contains 3 white and 5 black balls. 4 balls from U_1 are transferred to an empty urn U_2. 2 balls are drawn from U_2 and they both happen to be white. What is the probability that the third ball taken from the same urn will be white? (Uspensky, 1937).

(i) Suppose the two balls drawn from U_2 are returned before the third ball is drawn.

The balls drawn from U_1 could be

$B_1 : 2w, 2b$

$B_2 : 3w, 1b$

* With kind permission of publishers McGraw-Hill

Let A be the event that the 2 balls drawn from U_2 were both whites, and C be the event that the third ball from U_2 is white.

$$P(B_1) = \frac{\binom{3}{2}\binom{5}{2}}{\binom{8}{4}} = \frac{3}{7};$$

$$P(B_2) = \frac{\binom{3}{3}\binom{5}{1}}{\binom{8}{4}} = \frac{1}{14};$$

$$P(A/B_1) = \frac{\binom{2}{2}}{\binom{4}{2}} = \frac{1}{6};$$

$$P(A/B_2) = \frac{\binom{3}{2}}{\binom{4}{2}} = \frac{1}{2}.$$

$$P(C/A \cap B_1) = P(C/B_1) = \frac{1}{2}$$

$$P(C/A \cap B_2) = P(C/B_2) = \frac{3}{4}$$

Thus, $$P(C/A) = \frac{\frac{3}{7}\cdot\frac{1}{6}\cdot\frac{1}{2}+\frac{1}{14}\cdot\frac{1}{2}\cdot\frac{3}{4}}{\frac{3}{7}\cdot\frac{1}{6}+\frac{1}{14}\cdot\frac{1}{2}} = \frac{7}{12}.$$

(ii) Suppose the two balls drawn from U_2 were not replaced.

We then have $P(C/A \cap B_1) = 0$

$P(C/A \cap B_2) = \frac{1}{2}$ and accordingly

$$P(C/A) = \frac{\frac{1}{14} \cdot \frac{1}{2} \cdot \frac{1}{2}}{\frac{3}{7} \cdot \frac{1}{6} + \frac{1}{14} \cdot \frac{1}{2}} = \frac{1}{6}.$$

1.4 PRIOR AND POSTERIOR DISTRIBUTIONS

Let $f(x|\theta)$ be the probability density function (pdf) of a random variable X.

In Bayesian framework θ is a random variable and the data x is given or fixed.

Bayes' Theorem,

$$P(B|A) = \frac{P(B)\,P(A/B)}{P(A)}$$

states that the distribution of a parameter θ given the data x is given by

$$\prod(\theta|x) = \frac{g(\theta)\,f(x|\theta)}{h(x)}$$

$$= c\,g(\theta)\,f(x|\theta) \qquad (1.2)$$

where,

$$c^{-1} = \int_{\Omega} \Pi(\theta|x)\,d\theta$$

and Ω is the parameter space of θ . $g(\theta)$ is the prior distribution and $\Pi(\theta|x)$ is the posterior distribution of θ.

Let $\underline{x} = (x_1, x_2, \ldots, x_n)$ be a random, sample from the density function $f(x|\theta)$. Given the data $\underline{x}$, from (1.2)

$$\prod(\theta|\underline{x}) = c\,g(\theta)\,L(\underline{x}|\theta) \qquad (1.3)$$

We interpret (1.3) as follows:

To start with we have the pre-sample or prior information about θ summarized by the prior distribution $g(\theta)$. Then we proceed to collect the data $\underline{x}$ and the corresponding likelihood function $L(\underline{x}|\theta)$ gives us additional information about θ. Using the famous Theorem of Bayes we combine the prior information and the information contained in the sample, and calculate our new or revised degree of belief about θ given by the posterior distribution,

$$\prod(\theta \mid \underline{x}) \propto g(\theta)\, L(\underline{x} \mid \theta)$$

which contains all the information about θ.

The process continues as additional information becomes available.

Let $L_1(\underline{x} \mid \theta)$ be the likelihood of a sample $\underline{x} = (x_1, x_2, \cdots, x_n)$. Our revised degree of belief about θ, based on the sample $\underline{x}$ is represented by the posterior distribution,

$$\prod_1(\theta \mid \underline{x}) \propto g(\theta)\, L_1(\underline{x} \mid \theta).$$

let $\underline{y} = (y_1, y_2, \cdots, y_m)$ be a second set of data independent of $\underline{x}$ and let the corresponding likelihood be $L_2(\underline{y} \mid \theta)$. We treat the posterior $\Pi_1(\theta \mid \underline{x})$ as the prior distribution for the data $\underline{y}$. Now the new revised degree of belief based on the combined samples $(\underline{x}, \underline{y})$ is given by the posterior,

$$\begin{aligned}\prod_2(\theta \mid \underline{x}, \underline{y}) &\propto \prod_1(\theta \mid \underline{x})\, L_2(\underline{y} \mid \theta)\\ &\propto g(\theta)\, L_1(\underline{x} \mid \theta)\, L_2(\underline{y} \mid \theta)\\ &\propto g(\theta)\, L_3(\underline{x}, \underline{y} \mid \theta)\end{aligned}$$

The result may be extended to subsequent independent samples. Thus, the posterior based on pooling the sample information is the same as the one based on successive revisions of degrees of beliefs.

Example 1.8:

A dealer received a large shipment of some electronic components. He had extensive experience on the quality of the product the vendor had supplied in the past. The dealer assigns the following probabilities $P(A_i)$ to the proportion of defectives A_i.

A_i	$P(A_i)$
0.02	0.40
0.04	0.30
0.06	0.15
0.08	0.10
0.10	0.05

(i) A random sample of 10 components was drawn from the lot and each item was carefully checked. Only one component was found defective. Revise the probability distribution of A_i in the light of this information.

Let n be the sample size and r be the number of defectives and let B be the event: $n = 10, \; r = 1$

A_i	$P(A_i)$	$P(B \mid A_i)$	$P(A_i)P(B \mid A_i)$	$P(A_i \mid B)$
0.02	0.40	0.16675	0.06670	0.25802
0.04	0.30	0.27701	0.08310	0.32146
0.06	0.15	0.34380	0.05157	0.19949
0.08	0.10	0.37773	0.03777	0.14610
0.10	0.05	0.38742	0.01937	0.07493
			$P(B) = 0.25851$	1.00000

(ii) Subsequently, a second sample of size 20 was drawn and 3 defectives were found in the sample. We now revise the revised probability distribution of A_i. Let C be the event: $n = 20, r = 3$. We now compute the posterior probabilities $P(A_i \mid C)$ using $P(A_i \mid B)$ as prior distribution.

A_i	$P(A_i)$	$P(C \mid A_i)$	$P(A_i)P(C / A_i)$	$P(A / C)$
0.02	0.25802	0.00647	0.00167	0.02551
0.04	0.32146	0.03645	0.01172	0.17904
0.06	0.19949	0.08601	0.01716	0.26215
0.08	0.14610	0.14144	0.02066	0.31561
0.10	0.07493	0.19012	0.01425	0.21769
			$P(C) = 0.06546$	1.00000

(iii) Suppose the two samples of 10 and 20 are pooled into a single sample of 30 components. After careful checking, 4 defectives were detected. Let D: $n = 30, r = 4$.

We will compute the posterior probabilities $P(A_i \mid D)$.

A_i	$P(A_i)$	$P(D \mid A_i)$	$P(A_i)P(D \mid A_i)$	$P(A_i \mid D)$
0.02	0.40	0.00259	0.00104	0.02557
0.04	0.30	0.02427	0.00728	0.17900
0.06	0.15	0.07108	0.01066	0.26211
0.08	0.10	0.12843	0.01284	0.31571
0.10	0.05	0.17707	0.00885	0.21761
			$P(D) = 0.04067$	1.00000

which is the same as $P(A_1|C)$ except for small rounding off errors.

The prior distribution is an essential component of Bayesian inference. There is no single answer to the question, "What should be the right prior?" For much of the time the prior information is subjective and is based on a person's own experience and judgement. Several solutions have been proposed by Savage (1962), Jeffreys (1961), Zellner (1971), Raiffa and Schlaifer (1961), Hartigan (1964) among others. We mention a few of the well-known approaches in the following:

Principle of Stable Estimation

The Principle of Stable Estimation (Savage, 1962) states that if the prior distribution is such that it is essentially constant over the region where the likelihood function is appreciably large and then gradually tapers off beyond this region, we may assume that prior is locally uniform and the posterior distribution may be approximately by the likelihood function.

In most practical problems the information contained in the sample will have much greater effect on the posterior inference than would the prior information one may have before the experiment is performed or the sample is drawn. The Principle of Stable Estimation is very useful in a situation where the likelihood dominates the prior.

(i) Uniform Prior

In a state of ignorance the prior distribution is accepted as being uniform. It appears that great minds like Gauss, Bernoulli and Laplace used this principle in some form or other in their work. It is claimed that Bayes himself used uniform prior in his revolutionary work, "An essay towards solving a problem of doctrine of chance", posthumously communicated to the Royal Society (Nov. 10, 1763) by his friend Richard Price (1723–1791). It should be noted that although Bayes used the Principle of Equal Distribution of Ignorance when little or no information is available about the event of interest, he later realized that this Principle may not be acceptable to future generations and so he discarded it.

The apparent success with uniform prior subscribed to the general idea that perhaps the uniform prior is the final answer. Jeffreys (1961) makes an interesting comment that there is no more need for such an idea than to suggest that an oven which cooked roast beef once cannot cook

anything other than roast beef. One should be cautious before invoking the uniform prior theory, for a careless and mechanical use of this principle may lead to contradiction and confusion. Suppose for example, we have very little prior information about θ and therefore, we adopt a uniform prior U_1. If we know little about θ, our information about a function of θ, say $U(\theta)$, will also be very vague and imprecise and hence, a uniform prior U_2 for $U(\theta)$ would be in order. However, the two priors U_1 and U_2 will not be uniform unless $U(\theta)$ is linear in θ (Guttman, 1970). For an excellent discussion on uniform prior, we refer the reader to Jeffreys' Theory of Probability (1961) Section 3.1.

(ii) Non-informative Prior

When we are in a state of ignorance about the parameter we need to choose a prior which will formally express our ignorance about the parameter. Such a prior is known as Non-informative and as the name suggests, it is a prior that contains no information about θ. If the prior is Non-informative, we should assign the same density to each $\theta \in \Omega$, which of course implies that prior $g(\theta)$ is uniform given by $g(\theta = K$, $\theta \in \Omega$.

The Non-informative prior often leads to a class of improper priors, improper in the sense that $\int_{\Omega} g(\theta)d\theta \neq 1$.

Jeffreys does not regard it as a serious challenge and contends that g (θ) is not intended to represent a probability distribution of θ. g (θ) is just a function of θ which express a person's degree of confidence about θ.

(iii) Jeffreys' Invariant Prior

One way to get around this contradiction and confusion is to adopt Jeffreys' invariant prior. Such a prior is known as non-informative, vague or ignorance prior. Jefferys suggested the following rules for choosing the non-informative prior $g(\theta)$:

A: If $\Omega = (-\infty),\infty$, assume θ is uniformly distributed or

$$g(\theta) = \text{constant}$$

B: If $\Omega = (0,\infty)$ assume log θ is uniformly distributed, i.e

$$g(\theta) \propto \frac{1}{\theta}.$$

Rule A is invariant under any linear transformation, $U = a\,\theta + b$.

Rule B is invariant under any exponential transformation, $U = \theta^k, k \neq 0$.

C: Rules A and B are members of a large family of priors $g(\theta) \propto |I(\theta)|^{\frac{1}{2}}$ where θ may be real or vector-valued parameter, and

$$I(\theta) = -E\left[\frac{\partial^2 \log f(x \mid \theta)}{\partial\theta_i\, \partial\theta_j}\right]$$

is Fisher information matrix.

Consider the following matrix.

(a) *Normal* (μ, σ^2) : σ *known* .

From Rule A, $g(\mu) =$ constant .

(b) *Normal* (μ, σ^2) : μ *known* .

From Rule B, $g(\sigma) \propto \dfrac{1}{\sigma}$

(c) *Normal* (μ, σ^2) : μ, σ *both unknown*

$$g(\mu, \sigma) = g_1(\mu) g_2(\sigma \mid \mu) .$$

Since any information we might have about the value of μ is not likely to change one degree of belief about σ, $g_2(\sigma \mid \mu) \approx g_2(\sigma)$ and

$$g(\mu, \sigma) \approx g_1(\mu) g_2(\sigma)$$

$$\propto \frac{1}{\sigma} \text{ from Rules } A \text{ and } B.$$

Thus, in the absence of any knowledge to the contrary, we assume that μ and σ are independently distributed.

(d) *Exponential*:

$$f(x \mid \theta) = \frac{1}{\theta} \exp\left(-\frac{x}{\theta}\right), \quad x, \theta > 0$$

From Rule B, $g(\theta) \propto \dfrac{1}{\theta}$.

(e) *Weibull*:

$$f(x \mid \theta, p) = \frac{p}{\theta}\, x^{p-1} \exp\left(-\frac{x^p}{\theta}\right), \qquad x, p, \theta > .$$

$$g(p, \theta) = g_1(p)\, g_2(\theta \mid p)$$

$$\approx g_1(p)\, g_2(\theta)$$

$$\propto \frac{1}{p\,\theta} \qquad \text{from Rules } A \text{ and } B.$$

Let us obtain the priors (a) – (e) using Rule C.

$$\text{(a) } f(x \mid \mu, \sigma) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \qquad -\infty < \mu \infty, \sigma > 0.$$

Assume σ is known.

$$\log f = C - \log\sigma - \frac{(x-\mu)^2}{2\sigma^2}$$

$$\frac{\partial}{\partial\mu} \log f = \frac{(x-\mu)}{\sigma^2}$$

$$\frac{\partial^2}{\partial\mu^2} \log f = -\frac{1}{\sigma^2}$$

$$I(\mu) = -E\left(\frac{\partial^2}{\partial\mu^2} \log f\right) = \frac{1}{\sigma^2}$$

$$g(\mu) \propto \{I(\mu)\}^{1/2} \propto \frac{1}{\sigma}$$

$$g(\mu) = \text{constant since } \sigma \text{ is known}.$$

(b) Assume μ is known.

$$\frac{\partial}{\partial\sigma} \log f = -\frac{1}{\sigma} + \frac{(x-\mu)^2}{\sigma^3}$$

$$-\frac{\partial^2}{\partial\sigma^2} \log f = -\frac{1}{\sigma^2} + \frac{3(x-\mu)^2}{\sigma^4}$$

$$I(\sigma) = -E\left(\frac{\partial^2}{\partial\sigma^2} \log f\right) = \frac{2}{\sigma^2} \propto \frac{1}{\sigma^2}$$

$$g(\sigma) \propto \frac{1}{\sigma}.$$

(c) Assume both μ and σ are unknown.

$$\frac{\partial^2}{\partial \mu^2} \log f = -\frac{1}{\sigma^2}$$

$$\frac{\partial^2}{\partial \mu \, \partial \sigma} \log f = \frac{-2(x-\mu)}{\sigma^3}$$

$$\frac{\partial^2}{\partial \sigma^2} \log f = \frac{1}{\sigma^2} - \frac{3\,(x-\mu)^2}{\sigma^4}$$

$$I(\mu, \sigma) = -E \begin{bmatrix} \dfrac{2}{\sigma^2} & 0 \\ 0 & \dfrac{1}{\sigma^2} \end{bmatrix}$$

$$|\,I\,(\mu, \sigma)\,| = \frac{2}{\sigma^4}$$

$$g\,(\mu, \sigma) \propto \frac{1}{\sigma^2}$$

which is different from $g(\mu, \sigma) \propto \dfrac{1}{\sigma}$ obtained by using Rules A. Box and Tiao (1973) remark that the extra factor, $\dfrac{1}{\sigma}$ arises due to ignoring the prior independence between μ and σ.

(d)
$$f(x \mid \theta) = \frac{1}{\theta} \exp\left(-\frac{x}{\theta}\right)$$

$$\log f = -\log \theta - \frac{x}{\theta}$$

$$\frac{\partial}{\partial \theta} \log f = -\frac{1}{\theta} + \frac{x}{\theta^2}$$

$$\frac{\partial^2}{\partial \theta^2} \log f = \frac{1}{\theta^2} - \frac{2x}{\theta^3}$$

$$-E\left(\frac{\partial^2}{\partial \theta^2} \log f\right) = \frac{1}{\theta^2}$$

$$g\,(\theta) \propto \frac{1}{\theta}.$$

(e) $$f(x \mid \theta, p) = \frac{p}{\theta} x^{p-1} \exp\left(-\frac{x^p}{\theta}\right), \quad x, p, \theta > 0.$$

$$\log f = \log p - \log \theta + (p-1)\log x - \frac{x^p}{\theta}$$

$$\frac{\partial}{\partial \theta} \log f = -\frac{1}{\theta} + \frac{x^p}{\theta^2}$$

$$\frac{\partial^2}{\partial \theta^2} \log f = \frac{1}{\theta^2} - \frac{2x^p}{\theta^3}$$

$$E\left(\frac{\partial^2}{\partial \theta^2} \log f\right) = -\frac{1}{\theta^2}$$

$$\frac{\partial^2}{\partial \theta \, \partial p} \log f = \frac{x^p}{\theta^2} \log x$$

$$\frac{\partial^2}{\partial p} \log f = \frac{1}{p} + \log x - \frac{x^p}{\theta} \log x$$

$$\frac{\partial^2}{\partial p^2} \log f = -\frac{1}{p^2} - \frac{x^p}{\theta} (\log x)^2$$

Di- and tri-gamma functions are defined by

$$\Psi(n) = \frac{\partial}{\partial n} \Gamma(n) = \int_0^\infty \exp(-u)\, u^{n-1} \log u \, du;$$

$$\Psi'(n) = \frac{\partial^2}{\partial n^2} \Gamma(n) = \int_0^\infty \exp(-u)\, u^{n-1} (\log u)^2 \, du$$

where $\Psi(n) = \frac{1}{n-1} + \Psi(n-1)$ (Abramowitz and Stegun, 1964).

Let $\frac{x^p}{\theta} = u$. We have

$$\frac{p}{\theta} x^{p-1} dx = du, \quad \log x = \frac{1}{p}(\log u + \log\theta)$$

$$E\left(\frac{\partial^2}{\partial \theta \, \partial p} \log f\right) = \int_0^\infty \frac{x^p}{\theta^2} (\log x)\, f(x \mid \theta, p)\, dx$$

$$= \frac{1}{\theta p} \int_0^\infty u\,(\log u + \log\theta) \exp(-u)\,du$$

$$= \frac{1}{\theta p} [\log\theta + \Psi(2)],$$

and similarly,

$$E\left(\frac{\partial^2}{\partial p^2}\log f\right) = -\frac{1}{p^2}\left[1 + \int_0^\infty u\,(\log u + \log\theta)^2 \exp(-u)\,du\right]$$

$$= -\frac{1}{p^2}\left[1 + \Psi'(2) + 2\Psi(2)\log\theta + (\log\theta)^2\right]$$

$$|I(\theta, p)| = -E\begin{vmatrix} \dfrac{\partial^2 \log f}{\partial\theta^2} & \dfrac{\partial^2 \log f}{\partial\theta\,\partial p} \\ \dfrac{\partial^2 \log f}{\partial\theta\,\partial p} & \dfrac{\partial^2 \log f}{\partial p^2} \end{vmatrix}$$

$$= \begin{vmatrix} \dfrac{1}{\theta^2} & -\dfrac{1}{\theta p}\{\Psi(2) + \log\theta\} \\ -\dfrac{1}{\theta p}\{\Psi(2) + \log\theta\} & \dfrac{1}{p^2}\{1 + \Psi'(2) + 2\Psi(2)\log\theta + (\log\theta)^2\} \end{vmatrix}$$

$$= \frac{1}{\theta^2 p^2}\{1 + \Psi'(2) - \Psi^2(2)\}$$

$$\propto \frac{1}{\theta^2 p^2},$$

and hence,

$$g(\theta, p) \propto \frac{1}{\theta p}.$$

Consider the binomial distribution,

$$f(x \mid \theta) = \binom{n}{x}\theta^x (1-\theta)^{n-x}, \quad \theta < \theta < 1, \quad x = 0, 1, 2, \cdots, n.$$

Jeffreys (1961, Section 3.1) noted some inconsistencies in assigning uniform prior to a parameter with a finite range.

$$0 < \theta < 1 \text{ implies } 0 < \frac{\theta}{1-\theta} = \phi(\theta) < \infty$$

and from Rule B it follows that

$$g(\theta) \propto \left(\frac{1-\theta}{\theta}\right) \frac{1}{(1-\theta)^2} \propto \frac{1}{\theta(1-\theta)}$$ which is a non-informative prior.

This is the prior Haldane had proposed. Jeffreys remarks that Haldane's prior gives infinite density at limits and suggests a prior $g(\theta) \propto \frac{1}{\sqrt{\theta(1-\theta)}}$ which follows from Rule C:

$$\log f(x \mid \theta) = \text{constant} + x \log \theta + (n-x) \log (1-\theta)$$

$$\frac{\partial}{\partial \theta} \log f = \frac{x}{\theta} - \frac{n-x}{1-\theta}$$

$$\frac{\partial^2}{\partial \theta^2} \log f = -\frac{x}{\theta^2} - \frac{n-x}{(1-\theta)^2}$$

$$\mid I(\theta) \mid = -E\left(\frac{\partial^2 \log f}{\partial \theta^2}\right) = \frac{n\theta}{\theta^2} + \frac{n(1-\theta)}{(1-\theta)^2} = \frac{n}{\theta(1-\theta)}$$

Thus, $$g(\theta) \propto \mid I(\theta) \mid^{1/2} \propto \frac{1}{\sqrt{\theta(1-\theta)}}$$

(iv) Natural conjugate prior (NCP)

It will be helpful to discuss the function of sufficient statistics in classical and Bayesian Analysis before we introduce Natural Conjugate Prior (Raiffa and Schlaifer, 1961).

When we have considerable amount of data, $\underline{x} = (x_1, x_2, \cdots, x_n)$ from some family of distribution $f(x \mid \theta)$, for brevity and simplicity we may be interested in searching for some function (S) of the observations which will describe the basic characteristics of the distribution as effectively as the sample $\underline{x}$. Such a function, if it exists, is known as Sufficient Statistic. Sufficient statistic exists for a number of standard distributions. For example:

(i) if $X_i \sim N(\mu, \sigma^2)$ the sample mean $\overline{X}$ is a sufficient statistic and $\sim N(\mu, \frac{\sigma^2}{n})$;

(ii) if $X_i \sim$ Poisson with parameter λ, $Y = \sum_{i=1}^{n} X_i$ is a sufficient statistic and is distributed as Poisson with parameter $n\lambda$;

(iii) if X_i ~ gamma with parameter a, $Y = \sum_{i=1}^{n} X_i$ is a sufficient statistics with parameter na;

(iv) if X_i ~ Bernoulli with parameter p,

$$Y = \sum_{i=1}^{n} X_i \text{ is a sufficient statistic} \sim B(n, p).$$

In each case we observe that the statistical analysis based on the sufficient statistic will be as effective as the one based on the entire data set $\underline{x}$.

As in frequentist framework, sufficient statistic plays an important role in Bayesian interference in constructing a family of prior distributions known as Natural Conjugate Prior (NCP). The family of prior distributions $g(\theta)$, $\theta \in \Omega$, is called a natural conjugate family if the corresponding posterior distribution belongs to the same family as $g(\theta)$.

DeGroot (1970) has outlined a simple and elegant method of constructing a conjugate prior for a family of distributions $f(x \mid \theta)$ which admits a sufficient statistic.

Following DeGroot (1970), let $\underline{x} = (x_1, x_2, \cdots, x_n)$ be i.i.d. as $f(x \mid \theta)$ for which a sufficient statistic $T(x_1, x_2, \cdots, x_n) \equiv T$ exists. From Factorization Theorem, we have the likelihood function,

$$L(\underline{x} \mid \theta) = u(x_1, x_2, \cdots, x_n) \; \vartheta[T(x_1, x_2, \cdots, x_n), \theta] \propto \vartheta(T, \theta)$$

which implies that there exists a pdf $g(\theta \mid T) \equiv g(\theta)$ such that

$$g(\theta) \propto \vartheta(T, \theta), \quad \theta \in \Omega$$

Comparing $L(\underline{x} \mid \theta)$ and $g(\theta)$ it follows that there is an interesting relationship between the family of priors $g(\theta)$ and the likelihood $L(\underline{x} \mid \theta)$, viz., $g(\theta) \propto L(\underline{x} \mid \theta)$.

Thus, looking at the likelihood function we will have a fair idea about how to construct the prior $g(\theta) \equiv g(\theta \mid T)$. Such a prior is known as conjugate prior. For the proof that $g(\theta)$ is indeed conjugate (i.e. the posterior based on $g(T, \theta)$ will belong to the same family as $g(T, \theta)$) we refer to DeGroot (1970).

Consider the gamma pdf,

$$f(x \mid \theta) = \frac{1}{\theta^m \Gamma(m)} x^{m-1} \exp\left(-\frac{x}{\theta}\right), \qquad x, \theta, m > 0.$$

$$L(\underset{\sim}{x} \mid \theta) \propto \frac{1}{\theta^{mn}} \exp\left(-\frac{T}{\theta}\right), T = \sum_{i=1}^{n} x_i \text{ is sufficient.}$$

It follows that the NCP, $g(\theta) \propto \frac{1}{\theta^{b+1}} \exp\left(-\frac{a}{\theta}\right)$, $a, b, \theta > 0$, which is inverted gamma.

The posterior distribution,

$$\Pi(\theta \mid \underset{\sim}{x}) = K\,[\theta^{mn+b+1}]^{-1} \exp\left(-\frac{T+a}{\theta}\right)$$

$$= \frac{(T+a)^{mn+b}}{\theta^{mn+b+1}\,\Gamma(mn+b)} \exp\left\{-\frac{T+a}{\theta}\right\}$$

which is an inverted gamma.

Consider the Bernoulli distribution,

$$f(x \mid p) = p^x (1-p)^{1-x}, 0 < x < 1, 0 < p < 1.$$

$$L(\underset{\sim}{x} \mid p) \propto p^y (1-p)^{n-y}, y = \sum_{i=1}^{n} x_i$$

is sufficient. Hence, the NCP $g(p) \propto p^{a-1}(1-p)^{b-1}$, a beta distribution with parameters (a, b). The posterior,

$$\Pi(p \mid \underset{\sim}{x}) = K\,p^{y+a-1}\,(1-p)^{n-y+b-1}$$

$$= \frac{1}{B(y+a, n-y+b)}\,p^{y+a-1}\,(1-p)^{n-y+b-1}$$

which is a beta distribution with parameters $(y+a, n-y+b)$.

(v) Minimal Information Prior

Zellner (1971) used the information theoretic approach to define a minimal information prior. Let

$$I_x(\theta) = \int f(x \mid \theta) \log f(x \mid \theta)\, dx$$

be a measure of information in the pdf $f(x \mid \theta)$.

The prior average information is defined as

$$\bar{I}_x(\theta) = \int I_x(\theta)\, g(\theta)\, d\theta$$

where $g(\theta)$ is a proper prior density of θ.

$\int g(\theta) \log g(\theta)\, d\theta$ measures the information in our prior $g(\theta)$.

$$G = \bar{I}_x(\theta) - \int g(\theta).\ \log g(\theta)\, d\theta$$

$$= \int I_x(\theta)\, g(\theta)\, d\theta - \int g(\theta) \log g(\theta)\, d\theta \qquad (1.4)$$

is defined as a measure of gain in information. The minimal information prior is the one that maximizes G for varying $g(\theta)$ subject to the condition $\int g(\theta)\, d\theta = 1$.

Consider $X \sim N(1, \theta^2)$.

$$f(x \mid \theta) = \frac{1}{\sqrt{2\pi}\theta} \exp\left(-\frac{x^2}{2\theta^2}\right), \qquad -\infty < x < \infty, \theta > 0.$$

$$I_x(\theta) = \frac{-1}{\sqrt{2\pi}} \int \frac{1}{\theta} \exp\left(-\frac{x^2}{2\theta^2}\right). \left\{\frac{1}{2}(\log 2\pi + 2\log\theta) + \frac{x^2}{2\theta^2}\right] dx$$

$$= -\frac{1}{2}(\log 2\pi + 1) - \log\theta.$$

From (1.4)

$$G = -\frac{1}{2}(\log 2\pi + 1) - \int g(\theta) \log\theta\, d\theta - \int g(\theta) \log g(\theta)\, d\theta.$$

We have to maximize

$$U = G + \lambda\left[\int g(\theta)\, d\theta - 1\right]$$

$$= \text{Constant} - \int g(\theta) \log\theta\, d\theta - \int g(\theta) \log g(\theta)\, d\theta$$

$$+ \lambda\left[\int g(\theta)\, d\theta - 1\right]$$

where λ is Lagrange multiplier.

$$\frac{\partial U}{\partial g(\theta)} = 0 \text{ implies}$$

$$-\log\theta - \log g(\theta) - 1 + \lambda = 0$$

or

$$\log\{\theta g(\theta)\} = \text{const}$$

or $$g(\theta) \propto \frac{1}{\theta},$$

the same as Jeffreys' prior.

(vi) Asymptotically Locally Invariant Prior

Hartigan (1964) derived a family of prior densities to represent our ignorance about θ using invariance techniques similar to those suggested by Jeffreys (1946). He named this family asymptotically locally invariant (ALI) prior.

Let $f(x \mid \theta)$ be a family of probability density funcitons where $\theta = (\theta_1, \theta_2, \cdots \theta_x)$, and let

$$E\left[\left(\frac{\partial}{\partial \theta_i}\right) \log f(x \mid \theta)\right] = 0 \tag{1.5}$$

$$E\left[\left\{\left(\frac{\partial}{\partial \theta_i}\right) \log f(x \mid \theta)\right\} \left\{\left(\frac{\partial}{\partial \theta_j}\right) \log f(x \mid \theta)\right\}\right]$$

$$+ \left[E\left(\frac{\partial^2}{\partial \theta_i \partial \theta_j}\right) \log f(x \mid \theta)\right]$$

$$= 0, i = 1, 2, \cdots, k; \quad j = 1, 2, \cdots, k \tag{1.6}$$

Let a_{ij} be the (i, j)-th element of the inverse of the matrix with the (i, j)-th element.

$E\left[\left(\frac{\partial}{\partial \theta_i}\right)\left(\frac{\partial}{\partial \theta_j}\right) \log f(x \mid \theta)\right]$. Then the ALI prior density h is given by the solutions of the equation,

$$\frac{\partial}{\partial \theta_p} \log h(\theta) = -\sum_i \sum_j E\left[\left(\frac{\partial}{\partial \theta_i} \frac{\partial}{\partial \theta_p} \log f\right)\left(\frac{\partial}{\partial \theta_j} \log f\right)\right] a_{ij} \tag{1.7}$$

if such solutions exist (Hartigan, 1964).

The ALI priors are easy to derive for exponential family of distributions.

For $k = 1$, i.e. for a single parameter family, eqs. (1.5) – (1.7) reduce to

$$E(l_1) = 0 \quad E(l_1^2 + l_2) = 0$$

and,

$$\frac{\partial}{\partial\theta}\log h(\theta) = -\frac{E(l_l\, l_2)}{E(l_2)} \text{ where } l_l = \left(\frac{\partial^l}{\partial\theta^l}\right)\log f(x \mid \theta)$$

Consider the binomial probability distribution,

$$f(x \mid p) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0, 1, 2, \cdots n \cdot$$

$$\log f = \text{constant } + x\log p + (n-x)\log(1-p)$$

$$l_1 = \frac{\partial}{\partial p}\log f = \frac{x}{p} - \frac{n-x}{(1-p)} = \frac{x-np}{p(1-p)}$$

$$l_2 = \frac{\partial^2}{\partial p^2}\log f = -\frac{x}{p^2} - \frac{(n-x)}{(1-p)^2}$$

$$E(l_1) = 0$$

$$E(l_1^2 + l_2) = \frac{npq}{p^2 q^2} - \left[\frac{np}{p^2} + \frac{nq}{q^2}\right] = 0$$

$$-E(l_2) = \frac{n}{pq}$$

$$E(l_1\, l_2) = -\frac{1}{p^3 q^3} E[(X - np)\ \{X(q-p) + np^2\}]$$

$$= -\frac{1}{p^3 q^3}[(npq + n^2p^2 - n^2p^2)(q-p)]$$

$$= \frac{-n(q-p)}{p^2 q^2}.$$

ALI prior is given by the solution of the equation,

$$\frac{\partial}{\partial p}\log h(p) = -\frac{E(l_1 l_2)}{E(l_2)} = \frac{p-q}{pq} = \frac{1}{1-p} - \frac{1}{p}$$

$$\log h(p) \propto -[\log\{p(1-p)\}]$$

$$\propto \log\left[\frac{1}{p(1-p)}\right].$$

Hence, $h(p) \propto \dfrac{1}{p(1-p)}$ is the same as the prior suggested by Haldane.

We now derive the joint ALI prior $h(\mu, \sigma)$ for $N(\mu, \sigma^2)$.

$$f(x \mid \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}, \quad \begin{array}{c} -\infty < \mu < \infty, \\ -\infty < x < \infty, \\ \sigma > 0. \end{array}$$

$$\log f = C - \log \sigma - \frac{1}{2\sigma^2}(x-\mu)^2$$

$$\frac{\partial}{\partial \mu} \log f = \frac{x-u}{\sigma^2}$$

$$\frac{\partial}{\partial \sigma} \log f = -\frac{1}{\sigma} + \frac{1}{\sigma^3}(x-\mu)^2$$

$$\frac{\partial^2 \log f}{\partial \mu\, \partial \sigma} = -\frac{2}{\sigma^3}(x-\mu)$$

$$\left(\frac{\partial}{\partial \mu} \log f\right)\left(\frac{\partial}{\partial \sigma} \log f\right) = -\frac{x-\mu}{\sigma^3} + \frac{1}{\sigma^5}(x-\mu)^3.$$

We have,

$$E\left(\frac{\partial}{\partial \mu} \log f\right) = E\left(\frac{\partial}{\partial \sigma} \log f\right) = 0$$

$$E\left[\left(\frac{\partial}{\partial \mu} \log f\right)\left(\frac{\partial}{\partial \sigma} \log f\right)\right] + E\left[\frac{\partial^2}{\partial \mu\, \partial \sigma} \log f\right] = 0.$$

Thus, the equations (1.5) and (1.6) are satisfied.

$$\text{Let} \quad \mathbf{A} = \begin{bmatrix} E\left(\dfrac{\partial^2 \log f}{\partial \mu^2}\right) & E\left(\dfrac{\partial}{\partial \mu} \dfrac{\partial}{\partial \sigma} \log f\right) \\ E\left(\dfrac{\partial}{\partial \mu} \dfrac{\partial}{\partial \sigma} \log f\right) & E\left(\dfrac{\partial^2 \log f}{\partial \sigma^2}\right) \end{bmatrix}$$

$$= \begin{bmatrix} -\dfrac{1}{\sigma^2} & 0 \\ 0 & -\dfrac{2}{\sigma^2} \end{bmatrix}$$

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \mathbf{A}^{-1} = \begin{bmatrix} -\sigma^2 & 0 \\ 0 & -\dfrac{\sigma^2}{2} \end{bmatrix}$$

Substituting in (1.7)

$$\frac{\partial}{\partial\theta_p}\log h(\theta) = -E\left[\left(\frac{\partial}{\partial\mu}\frac{\partial}{\partial\theta_p}\log f\right)\left(\frac{\partial}{\partial\mu}\log f\right)\right]a_{11}$$

$$-E\left[\left(\frac{\partial}{\partial\sigma}\frac{\partial}{\partial\theta_p}\log f\right)\left(\frac{\partial}{\partial\sigma}\log f\right)\right]a_{22}$$

$$= E\left[\left(\frac{\partial}{\partial\mu}\frac{\partial}{\partial\theta_p}\log f\right)\cdot(X-\mu)\right]$$

$$+E\left[\left(\frac{\partial}{\partial\sigma}\frac{\partial}{\partial\theta_p}\log f\right)\left\{\frac{-\sigma}{2}+\frac{(X-\mu)^2}{2\sigma}\right\}\right]$$

where $\theta = (\mu, \sigma)$. For $p = 1$,

$$\frac{\partial}{\partial\mu}\log h(\mu,\sigma) = E\left[\left(\frac{\partial^2}{\partial\mu^2}\log f\right)(X-\mu)\right]$$

$$+E\left[\left(\frac{\partial^2}{\partial\sigma\,\partial\mu}\log f\right)\left\{-\frac{\sigma}{2}+\frac{(X-\mu^2)}{2\sigma}\right\}\right]$$

$$= 0$$

which implies $\log h(\mu,\sigma)$ is not a function of μ.

For $p = 2$,

$$\frac{\partial}{\partial\theta_p}\log h(\theta) = E\left[\left(\frac{\partial_2}{\partial\mu\,\partial\sigma}\log f\right)(x-\mu)\right]$$

$$+E\left[\left(\frac{\partial^2}{\partial\sigma^2}\log f\right)\left\{-\frac{\sigma}{2}+(X-\mu)^2\right\}\right]$$

$$= E\left[-\frac{2}{\sigma^3}(X-\mu)^2\right]$$

$$+E\left[\left\{\frac{1}{\sigma^2}-\frac{3}{\sigma^4}(X-\mu)^2\right\}\left\{-\frac{\sigma}{2}+\frac{(X-\mu^2)}{2\sigma}\right\}\right]$$

$$= -\frac{2}{\sigma}+E\left[-\frac{1}{2\sigma}+\frac{2}{\sigma^3}(X-\mu)^2-\frac{3(X-\mu)^4}{2\sigma^5}\right]$$

$$= -\frac{1}{2\sigma}-\frac{3}{2\sigma^5}E(X-\mu)^4 \tag{1.8}$$

$$E(X-\mu)^4 = \{E(X-\mu)^2\}^2+\mathrm{Var}(X-\mu)^2 .$$

$$= \sigma^4+\mathrm{Var}(X-\mu)^2$$

$\left(\frac{X-\mu}{\sigma}\right)^2$ is distributed as a χ^2 with 1 degree of freedom, so

$$\text{Var}\left(\frac{X-\mu}{\sigma}\right)^2 = 2 \text{ or } \text{Var}(X-\mu)^2 = 2\sigma^4$$

Substituting in eq. (1.8)

$$\frac{\partial}{\partial \sigma} \log h(\mu, \sigma) = -\frac{1}{2\sigma} - \frac{9}{2\sigma} = -\frac{5}{\sigma}.$$

$\log h(\mu, \sigma) \propto \log\left(\frac{1}{\sigma^5}\right)$, and hence,

$h(\mu, \sigma) \propto \frac{1}{\sigma^5}$ is the joint ALI prior for (μ, σ).

Hartigan (1964) pointed out that in some instances, the posterior distribution based on the ALI prior may lead to a chi-square having degrees of freedom contrary to the usual rule of assigning degrees of freedom to chi-square. For example, given a random sample from N (μ, σ^2) the ALI prior $(\mu, s) \frac{1}{\sigma^5}$ leads to a posterior distribution of σ, such that

$\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{\sigma^2}$ is distributed as a χ^2 with $(n+3)$ of freedom which is contrary to the well-known result that $\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{\sigma^2}$ is a χ^2 with $(n-1)$ degree of freedom.

(vii) Dirichlet's Prior

Dirichlet's prior distribution,

$$g(p_1, p_2, \cdots, p_k) = \frac{\Gamma(\theta)}{\Gamma(\theta_1)\,\Gamma(\theta_2)\cdots\Gamma(\theta_k)}\, p_1^{\theta_1 - 1}\, p_2^{\theta_2 - 1} \cdots p_k^{\theta_k - 1} \quad (1.9)$$

where $\quad \theta = \sum_{i=1}^{k} \theta_i, \ \sum_{i=1}^{k} p_i = 1, 0 < p_i < 1, \theta_i > 0$

is a generalization of the beta-prior,

$$g(p) = \frac{1}{B(a,b)} p^{a-1} (1-p)^{b-1}, 0 < p < 1, a, b > 0$$

and is conjugate to the multinomial family,

$$f(n_1, n_2, \cdots, n_k \mid p_1, p_2, \cdots, p_k) = \frac{n!}{n_1!n_2! \cdots n_k!}$$

$$(p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k}), \; n = \sum_{i=1}^{k} n_i. \qquad \text{(Wilks, 1962)}$$

Consider the prior expectation of P_i

$$E(P_i) = \int_0^1 \int_0^1 \cdots \int_0^1 p_i g(p_1, p_2 \cdots, p_i \cdots, p_k) \, dp_1 \, dp_2 \cdots dp_i \cdots dp_k$$

The evaluation of this multiple integral will be easy if we note that

$$\frac{\Gamma(\theta_1)\Gamma(\theta_2) \cdots \Gamma(\theta_i) \cdots \Gamma(\theta_k)}{\Gamma(\theta_1 + \theta_2 + \cdots + \theta_i \cdots + \theta_k)} = \int_0^1 \int_0^1 \cdots \int_0^1 p_1^{\theta_1 - 1}$$

$$p_2^{\theta_2 - 1} \cdots p_i^{\theta_i - 1} p_k^{\theta_k - 1} \, dp_1 \, dp_2 \cdots dp_i \cdots dp_k$$

which follows from eq. (1.9); thus

$$E(P_i) = \frac{\Gamma(\theta_1 + \theta_2 + \cdots + \theta_i + \cdots + \theta_k)}{\Gamma(\theta_1)\Gamma(\theta_2) \cdots \Gamma(\theta_i) \cdots \Gamma(\theta_k)} \int_0^1 \int_0^1 \cdots \int_0^1 p_1^{\theta_1 - 1}$$

$$p_2^{\theta_2 - 1} \cdots p_i^{(\theta + 1) - 1} \cdots p_k^{\theta_k - 1} \, dp_1 \, dp_2 \cdots dp_i \cdots dp_k$$

$$= \left[\frac{\Gamma(\theta_1 + \theta_2 + \cdots + \theta_i \cdots + \theta_k)}{\Gamma(\theta_1)\Gamma(\theta_2) \cdots \Gamma(\theta_i) \cdots \Gamma(\theta_k)}\right]$$

$$\left[\frac{\Gamma(\theta_1)\Gamma(\theta_2) \cdots \Gamma(\theta_i + 1) \cdots \Gamma(\theta_k)}{\Gamma(\theta_1 + \theta_2 + \cdots + \theta_i \cdots + \theta_k + 1}\right]$$

$$= \frac{\theta_i}{\theta} \qquad (10)$$

In many scientific investigations, the parameters $(p_1, p_2, \cdots, p_k)$ may not be *a-priori* independent. Dirichler's prior will be more appropriate in such a situation. See Section 2.10 for a case in point (Sinha and Guttman, 1994).

1.5 LOSS FUNCTIONS

Let θ be an unknown parameter of some distribution $f(x \mid \theta)$ and suppose we estimate θ by some statistic $T(x) \equiv T$. Let $L(T, \theta)$ represent the loss incurred when the true value of the parameter is θ and we are estimating θ by the statistic T.

Bayes estimator T^* is defined as the estimator that minimizes the posterior expected loss

$$E_{\theta \mid \underline{x}} [L(T^*, \theta)] = \int_{\Omega} L(T^*, \theta)\ \Pi\ (\theta \mid \underline{x})\, d\theta$$

where $\underline{x} = (x_1, x_2, \cdots, x_n)$ is the data and Ω is the parameter space.

We now consider the following loss functions:

(i) Linear Loss,

$$L(T, \theta) = c_1(T - \theta),\ T \geq \theta$$

and

$$= c_2(\theta - T),\ T < \theta$$

The constant c_1 and c_2 reflect the effect of over and under estimating θ.

By suitably choosing c_1, c_2 any fractile of the posterior distribution will be a Bayes estimator (Box and Tiao, 1973).

If c_1 and c_2 are functions of θ, the above loss function is called weighted linear loss function.

(ii) Absolute Error Loss,

$$L(T, \theta) = |T - \theta|$$

is called the absolute error loss function.

For such a loss function, Bayes estimator is the posterior median (DeGroot, 1970).

(iii) Zero-one Loss,

$$L(T, \theta) = 0 \quad \text{if } |T - \theta| \leq c$$

and, $\qquad = 1 \quad \text{if } |T-\theta| > c$

where c is a small positive constant. Bayes estimator is the mode of the posterior distribution (Raiffa and Schlaifer, 1961).

The Risk function $R(T,\theta)$ associated with the estimator T is defined as the expected value of the loss function.

The loss is zero if we made the correct decision about T and the loss is one if we made an incorrect decision about T.

$$\begin{aligned} R(T,\theta) &= E_x(L(T,\theta)] \\ &= \int L(T,\theta) f(x \mid \theta)\, dx \\ &= P[\,|T-\theta| > c] \\ &= P\text{ [incorrect decision about T]} \end{aligned}$$

which nearly implies that in the test of hypothesis context, the above risk represents the probability of Type I or Type II error (Berger, 1985).

(iv) Quadratic loss,

$$L(T,\theta) = K(T-\theta)^2$$

Such a loss function is widely used in most estimation problems.

Let is write $L(T,\theta) \equiv L(T)$.

Expanding $L(T)$ by Taylor's series,

$$\begin{aligned} L(T) &= L(\theta + T - \theta) \\ & L(\theta) + (T-\theta)\left[\frac{\partial L}{\partial T}\right]_{T=\theta} + \frac{(T-\theta)^2}{2\,!}\left[\frac{\partial^2 L}{\partial T^2}\right]_{T=\theta} \\ & + \frac{(T-\theta^3)}{3\,!}\left[\frac{\partial^3 L}{\partial T^3}\right]_{T=\theta} + \cdots \end{aligned}$$

we have $L(\theta) = 0$ and for Min $L(T)$, $\left[\frac{\partial L}{\partial T}\right] = 0$, $T = \theta$.

We are looking for an estimator T which, for all practical purposes, should be closed to θ and thereby we may ignore higher order terms like $(T-\theta)^r$, $r>2$ Hence, we may approximate as

$$L(T, \theta) \approx K(T - \theta)^2 .$$

If K is a function of θ, the loss function is called the weighted quadratic loss function. If $K = 1$, we have

$L(T, \theta) = (T - \theta)^2$, known as the Squared-Error-Loss Function (SELF).

Consider Bayes estimator under SELF

$$E_{\theta | \underline{x}}[L(T, \theta)] = \int_\Omega L(T, \theta)\, \Pi(\theta | \underline{x})\, d\theta .$$

For Min $\{E_{\theta | \underline{x}}[L(T, \theta)]\}$, $\dfrac{\partial}{\partial T} \displaystyle\int_\Omega L(T, \theta)\, \Pi(\theta | \underline{x})\, d\theta = 0$

or,
$$\frac{\partial}{\partial T} \int_\Omega (T - \theta)^2\, \Pi(\theta | \underline{x})\, d\theta = 0,$$

leading to Bayes estimator T^* which satisfies

$$\int_\Omega (T^* - \theta)\, \Pi(\theta | \underline{x})\, d\theta = 0$$

i.e.,
$$T^* = \int_\Omega \theta\, \Pi(\theta | \underline{x})\, d\theta$$

$$= \text{posterior mean} .$$

Thus, under SELF, Bayes estimator is the posterior expectation.

(Unless otherwise stated, we will be using SELF for our estimation problems).

(v) Linex Loss.

A symmetric loss function assumes that positive and negative errors are equally serious. However, in some estimation problems such an assumption may be inappropriate. A positive error may be more serious than a negative error or vice-versa (Ferguson, 1967; Zellner and Geisel, 1968; Achison and Dunsmore, 1975, Berger, 1985). Zellner (1986) derived and discussed the properties of Varian's (1975) asymmetric LINEX loss function for a number of distributions. Such a loss function is defined as

$$L(\Delta) = b \exp(A\, \Delta) - c\, \Delta - b, \; \Delta = T - \theta, \; a, c \neq 0, \; b > 0.$$

Note that $L(0) = 0$ and,

$$\frac{\partial}{\partial \Delta} L(\Delta) = ab \exp(a\Delta) - c.$$

For a minimum to exist at $\Delta = 0$,

$$\left[\frac{\partial}{\partial \Delta} L(\Delta)\right]_{\Delta=0} = 0 = ab - c$$

and we have a two-parameter loss function,

$$L(\Delta) = b[\exp(a\Delta) - a\Delta - 1] \quad b > 0, a > 0 \qquad (1.11)$$

1.6 BAYES RISK

Bayes risk associated with an estimate T is defined as the expected value of the risk function $R(T, \theta)$ with respect to the prior distribution $g(\theta)$ of θ, and is given by

$$\begin{aligned} R^*(T,\theta) &= E_\theta[R(T,\theta)] \\ &= \int R(T,\theta)\, g(\theta)\, d\theta \\ &= \int E_x[L(T,\theta)]\; g(\theta)\, d\theta \\ &= \int \left[\int L(T,\theta) f(x \mid \theta)\, dx\right] g(\theta)\; d\theta \end{aligned} \qquad (1.12)$$

We now obtain Bayes risks associated with Bayes estimators of the parameters of Binomial, Poisson, Gamma distributions under SELF and Normal distribution using Linex loss function.

Example 1.9:

Let $X \sim B(n, p)$ and the prior

$$g(p) \propto \frac{1}{\sqrt{p(1-p)}}, \quad 0 < p < 1. \qquad (1.13)$$

Derive Bayes estimator of p and obtain Bayes risk associated with the estimator $p^* = E(p \mid x)$.

$$f(x \mid p) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, 2, \cdots n$$

$$\Pi(p \mid x) \propto g(p)\, f(x \mid p) \propto p^{x+\frac{1}{2}-1} \, (1-p)^{n-x+\frac{1}{2}-1}$$

Normalizing the posterior we have

$$\Pi(p \mid x) = \frac{1}{B\left(x+\frac{1}{2}, n-x+\frac{1}{2}\right)} p^{x+\frac{1}{2}-1} (1-p)^{n-x+\frac{1}{2}-1}$$

$$p^* = E(p \mid x) = \frac{x+\frac{1}{2}}{n+1}$$

$$E_X(p^*) = \frac{np+\frac{1}{2}}{n+1}$$

$$V_X(p^*) = \frac{np(1-p)}{(n+1)^2}$$

$$R(p^*, p) = E_X(p^*-p)^2$$

$$= E_X\left[\frac{X+\frac{1}{2}}{n+1} - \frac{np+\frac{1}{2}}{n+1} + \frac{np+\frac{1}{2}}{n+1} - p\right]^2$$

$$= V_X(p^*) + \left(\frac{np+\frac{1}{2}}{n+1} - p\right)^2$$

$$= \frac{1}{(n+1)^2}\left[np(1-p) + \left(\frac{1}{2}-p\right)^2\right]$$

$$= \frac{1}{(n+1)^2}\left[\frac{1}{4} + p(n-1) - p^2(n-1)\right]. \tag{1.14}$$

From results (1.12) and (1.14), Bayes risk

$$R^*(p^*, p) = \int_0^1 E_X(p^*-p)^2 g(p)\, dp$$

$$= \frac{1}{(n+1)^2}\left[\frac{1}{4} + (n-1)E(p) - (n-1)E(p^2)\right] \tag{1.15}$$

From (1.13)

$$g(p) = \frac{1}{B\left(\frac{1}{2}, \frac{1}{2}\right)} p^{\frac{1}{2}-1} (1-p)^{\frac{1}{2}-1} \qquad 0 < p < 1$$

$$E(p) = \frac{1}{2}, \quad E(p^2) = \frac{3}{8}.$$

Substituting in (1.15)

$$R*(p*, p) = \frac{1}{(n+1)^2} \left[\frac{1}{4} - \frac{1}{2}(n-1) - \frac{3}{8}(n-1)\right]$$

$$= \frac{1}{8(n+1)}.$$

Example 1.10:

Consider the gamma density function,

$$f(x \mid \theta) = \frac{1}{\theta^n \Gamma(n)} x^{n-1} \exp\left(-\frac{x}{\theta}\right), x, \theta > 0$$

and the corresponding conjugate prior,

$$g(\theta) = \frac{\exp\left(-\frac{1}{\theta}\right)}{6\theta^5} \qquad \theta > 0.$$

Derive Bayes risk relative to θ*, the posterior expectation of θ.

The posterior density,

$$\Pi(\theta \mid x) = \frac{K}{\theta^{n+5}} \exp\left(-\frac{x+1}{\theta}\right)$$

$$= \frac{(x+1)^4 \exp\left(-\frac{x-1}{\theta}\right)}{\Gamma(n+4)\ \theta^{n+5}}$$

$$\theta^* = E(\theta \mid x) = \frac{x+1}{n+3}$$

$$R(\theta^*, \theta) = E_X(\theta^* - \theta)^2$$

$$= E_X\left[\theta^* - \frac{n\theta + 1}{n+3} + \frac{n\theta + 1}{n+3} - \theta\right]^2$$

$$= V_X(\theta^*) + \left(\frac{1 - 3\theta}{n+3}\right)^2$$

$$= \frac{n\,\theta^2 + (1 - 3\theta)^2}{(n+3)^2}$$

$$= \frac{(n+9)\,\theta^2 - 6\,\theta + 1}{(n+3)^2}$$

$$E(\theta) = \int_0^\infty \theta\, g(\theta)\, d\theta = \frac{1}{3}\cdot \text{ Similarly, } E(\theta^2) = \frac{1}{6}\cdot$$

Hence,

$$R^*(\theta^*, \theta) = \frac{1}{(n+3)^2}\left[(n+9)\,E(\theta^2) - 6\,E(\theta) + 1\right]$$

$$= \frac{1}{(n+3)^2}\left[\frac{n+9}{6} - 2 + 1\right] = \frac{1}{6(n+3)}\cdot$$

Example 1.11:

Let $(x_1, x_2, \ldots, x_n)$ be i.i.d. as

$$f(x\,|\,\lambda) = \frac{\lambda^x \exp(-\lambda)}{x\,!}, \quad x = 0, 1, 2, \cdots$$

where λ is distributed as gamma with parameters α and β.

Given $\underline{x} = (x_1, x_2, \cdots, x_n)$ obtain Bayes estimator of λ and the associated Bayes risk.

We have $L(\underline{x}\,|\,\lambda) \propto \exp(-n\lambda)\,\lambda^y, y = \sum_{i=1}^{n} x_i$

and $$g(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)}\,\lambda^{\alpha-1}\exp(-\beta\lambda), \quad \lambda, \alpha, \beta > 0\,.$$

$$\Pi(\lambda\,|\,\underline{x}) = \frac{(\beta+n)^{y+\alpha}}{\Gamma(y+\alpha)}\,\lambda^{y+\alpha-1}\exp\{-(n+\beta)\,\lambda\}$$

Bayes estimator $\lambda^* = E(\lambda \mid \underline{x}) = \dfrac{y + \alpha}{n + \beta}$

$$E_x(\lambda^*) = \frac{n\lambda + \alpha}{(n+\beta)}, \; V_x(\lambda^*) = \frac{n\lambda}{(n+\beta)^2}$$

$$R(\lambda^*, \lambda) = E_x(\lambda^* - \lambda)^2$$

$$= V_x(\lambda^*) + \left(\frac{n\lambda + \alpha}{n+\beta} - \lambda\right)^2$$

$$= \frac{n\lambda + (\alpha - \lambda\beta)^2}{(n+\beta)^2}$$

$$= \frac{\alpha^2 + \lambda(n - 2\alpha\beta) + \lambda^2\beta^2}{(n+\beta)^2}.$$

From the prior distribution $g(\lambda)$, we obtain

$$E(\lambda) = \frac{\alpha}{\beta}, \quad E(\lambda^2) = \frac{\alpha(\alpha+1)}{\beta^2}.$$

Hence, from (1.12) we have Bayes risk

$$R^*(\lambda^*, \lambda) = \int_0^\infty E_\lambda(\lambda^* - \lambda)^2 g(\lambda)\, d\lambda$$

$$= \frac{1}{(n+\beta)^2}\left[\alpha^2 + (n - 2\alpha\beta)E(\lambda) + \beta^2 E(\lambda^2)\right]$$

$$= \frac{1}{(n+\beta)^2}\left[\alpha^2 + (n - 2\alpha\beta)\frac{\alpha}{\beta} + \alpha(\alpha+1)\right]$$

$$= \frac{\alpha}{\beta(n+\beta)}.$$

Example 1.12:

Let $\quad X \sim N(\mu, 1)$

$\quad \mu \sim N(0, 1).$

Show that relative to the LINEX loss function

$$L(\Delta) = b\,[\exp(a\Delta) - a\Delta - 1], \Delta = T - \mu, b > 0, a \neq 0,$$

Bayes estimator of μ is given by

$\mu^* = \dfrac{2x-a}{4}$ and the associated Bayes risk $R^*(\mu^* \mid \mu) = \dfrac{a^2 b}{4}$.

$$E_\mu [L(\Delta)] = E_\mu [L(T-\mu)]$$

$$= b\,[\exp(a\,T) E_\mu \{\exp(-a\,\mu)\} - a\,\{T - E(\mu)\} - 1].$$

Bayes estimator T^* is the solution of the equation,

$$\frac{\partial}{\partial T}\left[E_\mu \{L(T-\mu)\}\right] = 0 = a \exp(aT^*) E_\mu \{\exp(-a\,\mu)\} - a$$

$$T^* = -\frac{1}{a}\log[E_\mu \{\exp(-a\,\mu)\}] \tag{1.16}$$

$$f(x \mid \mu) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2}\right\}$$

$$g(\mu) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\mu^2}{2}\right).$$

The posterior distribution,

$$\Pi(\mu \mid x) = K \exp\left[-\frac{1}{2}\{(x-\mu)^2 + \mu^2\}\right]$$

$$= \frac{1}{\sqrt{\pi}} \exp\left[-\left(\mu - \frac{x}{2}\right)^2\right]$$

$$E_\mu \{\exp - (-a\,\mu)\} = \int_{-\infty}^{\infty} \exp(-a\,\mu)\, \Pi(\mu \mid x)\, d\mu$$

$$= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \exp - \left\{a\,\mu + \left(\mu - \frac{x}{2}\right)^2\right\} d\mu$$

$$= \exp\left\{\frac{(x-a)^2 - x^2}{4}\right\}$$

$$= \exp\left[-\frac{2x\,a - a^2}{4}\right].$$

From (1.16)

$$\mu^* = \frac{2x-a}{4}.$$

Note that $X \sim N(\mu, 1)$ and $\mu \sim N(0, 1)$.

Hence, the moment generating function (mgf) of X and μ are exp $\left(\mu t + \frac{t^2}{2}\right)$ and $\exp\left(\frac{t^2}{2}\right)$ respectively.

$$E_X \{L(\Delta)\} = b\,[\exp(a\mu^{*})\exp(-a\mu) - a(\mu^* - \mu) - 1]$$

$$= b\left[\exp\left(-a\mu - \frac{a^2}{4}\right) E_X\left\{\exp\left(\frac{aX}{2}\right)\right\} + a\left(\frac{a}{4} + \mu\right) - \frac{1}{2}E(X) - 1\right]$$

$$= b\left[\exp\left(-a\mu - \frac{a^2}{4}\right)\left\{\text{mgf of } \left(\frac{aX}{2}\right)\right\} + \frac{a^2}{4} + \frac{a\mu}{2} - 1\right]$$

$$= b\left[\exp\left(-a\mu - \frac{a^2}{4} + \frac{a\mu}{2} + \frac{a^2}{8}\right) + \frac{a^2}{4} + \frac{a\mu}{2} - 1\right]$$

Hence, Bayes risk

$$R^*(\mu^*, \mu) = b\left[\exp\left(\frac{-a^2}{8}\right) E_\mu\left\{\exp\left(\frac{-a\mu}{2}\right)\right\} + \frac{a^2}{4} - 1\right]$$

$$= b\left[\exp\left(-\frac{a^2}{8}\right)\left\{\text{mgf of } \left(-\frac{a\mu}{2}\right)\right\} + \frac{a^2}{4} - 1\right]$$

$$= b\left[\exp\left(-\frac{a^2}{8} + \frac{a^2}{8}\right) + \frac{a^2}{4} - 1\right]$$

$$= \frac{a^2 b}{4}.$$

1.7 BAYESIAN VS CLASSICAL — WHY BAYESIAN ?

Bayesian approach to statistics is very different from the classical methodology and the difference may be traced in the way the two schools interpret probability. As defined earlier, the classical or frequentists interpret probability as the limit of the success ratio as the number of

trials n conceptually tends to infinity. Under this interpretation the parameter θ in a statistical model is treated as an unknown constant and the sample of observations, $\underline{x} = (x_1, x_2, x_3, \cdots, x_n)$ is regarded as a random sample from some underlying distribution $F(x \mid \theta)$. The classical school believes in Fisher's likelihood Principle which claims that all the information about the unknown parameter(s) is contained in the sample as summarized by the likelihood function. This principle leads to Fisher's maximum likelihood estimator. Notwithstanding certain limitations, the maximum likelihood estimators have a number of desirable properties and are extensively used in preference to other classical estimators.

A Bayesian, however, interprets probability as a person's degree of belief in a certain proposition A based on the prior (or current) knowledge about A and this '*degree of belief*' is successively revised or updated as new information is accumulated about the proposition.

In Bayesian framework the parameter is justifiably regarded as a random variable and the data once obtained, is given or fixed.

Thus, the basic difference in the two approaches may be explained in the single sentence that to a frequentist, *the parameter is constant and he is suspicious about the data*, whereas to a Bayesian, *the data is given (or fixed) and he is suspicious about the parameter.*

Fisher (1936) remarked that Bayes' Theorem was the first attempt known to us to rationalize the process of inductive reasoning and bring inductive inference within the domain of the theory of probability. Prior to Bayes, sampling theorists developed several methods whereby given a population, one could state the probability that a particular sample would be obtained. Bayes' problem was: given a sample, what is the probability that the sample was drawn from a particular population? (Kenny and Keeping, 1951). Bayes Theorem attempts to answer this question.

Bayes Theorem is not just a curious speculation in the doctrine of chance. It provides a sound foundation for all our reasonings concerning past facts and what is likely to be in future and it provides a measure of the probability with which past experiences might be applied to events hereafter. In recent past Bayesian approach to statistics was looked down upon as being controversial and even somewhat suspicious. It has now come

to be recognized as an alternative and interesting approach that deserves serious study of its applied implications and theoretical foundation.

Bayesian framework has several interesting features that make it more attractive to applied statisticians than to its frequentist counterpart. It combines the prior information with the information contained in the data to formulate the posterior distribution which is the basis of Bayesian inference. It possesses a well-developed and straight forward procedure for facing the problem of optimal action in a state of uncertainty. Application of Bayesian concepts and methods abound in Econometrics, Sociology, Engineering, Reliability estimation and Quality control. It addresses the question of how the model underlying the data may be revised in the light of new information and experience.

If the parameter is non-negative, Bayes estimator will be non-negative. This avoids the possibility of an essentially non-negative parameter having a negative estimator which could occur when estimating using sampling framework. Given the sample and the corresponding likelihood function, Bayes estimator will be unique and there is no confusion about which estimator or estimators should be chosen, unlike the estimator(s) sometimes obtained via classical methodology.

It is very useful in a situation where the underlying distribution depends on, say, r number of parameters $\theta = (\theta_1, \theta_2, \cdots, \theta_k, \theta_{k+1}, \cdots, \theta_r)$, and we are interested only in the subset $\alpha = (\theta_1, \theta_2, \theta_3, \cdots, \theta_k)$. The subset $\beta = (\theta_{k+1}, \theta_{k+2}, \cdots \theta_r)$ is known as nuisance parameters. While in non-Bayesian case, one would face considerable difficulties in dealing with nuisance parameters, in Bayesian setup inference on the subset α may be based upon the marginal posterior of α obtained by integrating out β from the posterior of θ (Guttman, 1970).

Perhaps the above introductory remarks will provide a modest reply to, "Why Bayesian"

EXERCISES

1. A large community in which the incidence of cancer is approximately 1 in 2000, is subjected to a cancer-screening test.

In 99% of the cases the test gives a positive reading when the person has the disease and in 98% of the case it gives a negative reading when the person is healthy. What is the probability that a person whose test shows a positive reading does not have cancer?

*2 From an urn containing 5 white and 5 black balls, 5 balls are transferred into an empty second urn. From there, 3 balls are transferred into an empty third urn and, finally, one ball is drawn from the latter. It turns out to be white. What is the probability that all 5 balls transferred from the first urn were white? (Uspensky, 1937).

3. A random variable X has the probability density function (pdf)

$$f(x \mid \theta) = \frac{\theta}{2^{\theta}} x^{\theta - 1}, \qquad 0 < x < 2 .$$

Derive Jeffreys invariant prior $g(\theta)$ and show that the corresponding estimator

$$\theta^* = \frac{n}{\log K} \quad \text{where } K = \prod_{i=1}^{n} \left(\frac{1}{x_i} \right).$$

4. Refer to the pdf in Exercise 3.

Show that the conditions (1.5) and (1.6) are satisfied, and ALI prior is given by $h(\theta)$ = constant. What is the corresponding estimate θ^*?

5. Show that the MIP, NCP and Jeffreys invariant priors $g(\beta)$ for the pareto distribution

$$f(x \mid \theta, \beta) = \frac{\beta\, \theta^{\beta}}{x^{\beta + 1}}, \quad 0 < \theta < x < \infty, \beta > 0,$$

where θ is known, are given by

$$g_1(\beta) \propto \beta \exp\left(-\frac{1}{\beta} \right),$$

$$g_2(\beta) \propto \beta^{a-1} \exp\left(-\frac{b}{\beta} \right), \qquad \text{and}$$

* With kind permission of publishers McGraw-Hill.

$$g_3(\beta) \propto \frac{1}{\beta}, \text{ respectively}.$$

6. Let $f(x \mid \theta) = \dfrac{1}{\theta}, \; 0 < x < \theta$

and the prior $g(\theta) = \theta \exp(-\theta)$.

Obtain, (i) Bayes estimator $\theta^* = E(\theta \mid x)$, and

(ii) Bayes risk $R^*(\theta^*, \theta)$.

7. Let $X \sim B(n, p)$ and

$$g(p) \propto p^{a-1}(1-p)^{b-1}, \; 0 < p < 1, \; a, b > 0.$$

Show that Bayes risk for the estimator p^* is

$$R^*(p^*, p) = \frac{a\,b}{(n+a+b)\,(a+b)\,(a+b+1)}.$$

8. Let $f(x \mid \theta) = \dfrac{1}{\Gamma(n)\,\theta^n}\, x^{n-1} \exp\left(-\dfrac{x}{\theta}\right), \quad x, \theta > 0.$

Show that under the NCP

$$g(\theta) \propto \frac{\exp\left(-\dfrac{a}{\theta}\right)}{\theta^{m+1}}, \quad a, m > 0$$

Bayes risk for the estimator θ^* is given by

$$R^*(\theta^*, \theta) = \frac{a^2}{(n+m-1)\,(m-1)\,(m-2)}, \; m > 2.$$

9. Derive MIP $g(\theta, p)$ for the Weibull pdf

$$f(x \mid \theta, p) = \frac{p}{\theta}\, x^{p-1} \exp\left(-\frac{x^p}{\theta}\right), \quad x, p, \theta > 0.$$

(Sinha and Zellner, 1990).

10. Are the conditions (1.5) and (1.6) for the ALI-prior satisfied for the Weibull pdf in Exercise 9?

11. Let $\underline{x} = (x_1, x_2, \cdots, x_n)$ be a random sample $N(\mu, \sigma^2)$, σ known, and let the prior $g(\mu) =$ constant. Show that

under the LINEX loss function discussed in Section 1.5, Bayes estimator $\mu^* = \bar{x} - \dfrac{a\,\sigma^2}{2n}$.

(Zellner, 1986)

12. Let $X \sim N(\mu, \sigma^2)$, σ known and $\mu \sim N(0, \sigma^2)$, $\underset{\sim}{x} = (x_1, x_2, \cdots, x_n)$.

Show that under the LINEX loss function, Bayes estimator

$$\mu^* = \frac{\bar{x}}{1+\lambda} - \frac{a\,\sigma^2}{2n(1+\lambda)} \quad \text{where} \quad \lambda = \frac{\sigma^2}{n}$$

and Bayes risk

$$R^*(\mu^* \mid \mu) = \frac{a^2\, b\, \sigma^2}{2n(1+\lambda)}.$$

(Zellner, 1986)

13. Refer to the data in Example 1.8 and let

A_i :	0.01	0.03	0.05	0.07	0.09
$P(A_i)$:	0.60	0.20	0.15	0.03	0.02

Compute the posterior probabilities in sections (i), (ii) and (iii).

14. Show that under the general loss function (GLF),

$$L(T, \theta) = \lambda(\theta)\,(T^m - \theta^m)^2, \; m > 0$$

Bayes estimator

$$T^* = \left[\frac{E\{\lambda(\theta)\,\theta^m \mid x\}}{E\{\lambda(\theta) \mid x\}}\right]^{\frac{1}{m}}, \quad \text{and}$$

as $m \to 0$, $\log T^* \to E(\log \theta)$

(El-Sayyad, 1967)

15. Show that the normalizing constant K in the joint distribution

$$g(p_1, p_2, p_3) = K\, p_1^{\theta_1 - 1}\, p_2^{\theta_2 - 1}\, p_3^{\theta_3 - 1}\,(1 - p_1 - p_2 - p_3)^{\theta_4 - 1},$$

$$0 < p_1 + p_2 + p_3 < 1$$

is given by

$$K = \frac{\Gamma(\theta_1 + \theta_2 + \theta_3 + \theta_4)}{\Gamma(\theta_1)\,\Gamma(\theta_2)\,\Gamma(\theta_3)\,\Gamma(\theta_4)} .$$

16. A letter (a closed enveloped) is believed to have come from either Allahabad or Jacobabad. The postal imprint is partially smudged except for two consecutive letter *ba* being clearly legible. What is the probability that the letter is coming from Jacobabad?

17. Given the data $\underline{x} = (x_1, x_2, \cdots, x_n)$ and the sufficient statistic $y = \sum_{i=1}^{n} x_i$, construct the NCP for the geometric distribution,

$$f(x \mid p) = p\,q^{x-1}, q = 1-p, \quad x = 1, 2, \cdots$$

Hence, obtain Bayes estimator p^*. When does p^* identifies itself with the MLE $\hat{p}$?

18. Given the data $\underline{x}$ and the sufficient statistic $y = \sum_{i=1}^{n} x_i$, construct the NCP for the Poisson distribution,

$$f(x \mid \lambda) = \exp(-\lambda)\frac{\lambda^x}{x!}, \; x = 0, 1, 2, \cdots$$

When does the Bayes estimator λ^* indentifies itself with the corresponding maximum likelihood estimator $\hat{\lambda}$?

CHAPTER

2 Bayesian Estimation of Parameters of some Well-known Distributions

2.1 NORMAL DISTRIBUTION

Let $\underset{\sim}{x}$ $(x_1, x_2, \ldots x_n)$ be a random sample of size n from a normal distribution $N(\mu, \sigma^2)$. The probability density function (pdf) of X is given by

$$f(x \mid \mu, \sigma) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}, \quad -\infty < x,\ \mu < \infty,\ \sigma > 0 \tag{2.1}$$

The likelihood function

$$L(\mu, \sigma \mid \underset{\sim}{x}) \propto \frac{1}{\sigma^n} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2\right\} \tag{2.2}$$

We will derive Bayes estimator of μ under the following situations:

(i) σ known: conjugate prior for μ

We may write the likelihood function as

$$L(\mu, \sigma \mid \underset{\sim}{x}) = \frac{1}{(\sqrt{2\pi}\,\sigma)^n} \exp\left[-\frac{1}{2\sigma^2}\{A + n(\bar{x} - \mu)^2\}\right]$$

$$= \exp\left[-\frac{n}{2\sigma^2}(\bar{x} - \mu)^2\right]\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \exp\left(-\frac{A}{2\sigma^2}\right)$$

where, $$A = \sum_{i=1}^{n} (x_i - \bar{x})^2.$$

It is well-known that $\bar{x}$ is sufficient.

$$K(\bar{x}, \mu) = \text{kernel of the likelihood function}$$

$$= \exp\left[-\frac{n}{2\sigma^2} (\bar{x} - \mu)^2\right].$$

Replacing $\bar{x}$ by m and $\frac{\sigma^2}{n}$ by δ^2, the natural conjugate prior

$$g(\mu) \propto \exp\left[-\frac{(\mu - m)^2}{2\delta^2}\right]$$

$$\propto N(m, \delta^2).$$

To normalize the prior we will use the result

$$\int_{-\infty}^{\infty} \exp(-\alpha x^2)\, dx = \frac{\sqrt{\pi}}{\sqrt{\alpha}}. \quad \text{(Appendix A.2)}$$

$$g(\mu) = K_1 \exp\left[-\frac{(\mu - m)^2}{2\delta^2}\right] d\mu$$

$$K_1^{-1} = \int_{-\infty}^{\infty} \exp\left[-\frac{(\mu - m)^2}{2d^2}\right] d\mu$$

$$= \sqrt{2\pi}\, \delta.$$

Combining the likelihood and the prior, the posterior distribution,

$$\prod(\mu \mid \underset{\sim}{x}) \propto \exp\left[\left\{\mu^2\left(\frac{n}{\sigma^2} + \frac{1}{\delta^2}\right\} - 2\mu\left(\frac{n\bar{x}}{\sigma^2} + \frac{m}{\delta^2}\right) + \left(\frac{n\bar{x}^2}{\sigma^2} + \frac{m^2}{\delta^2}\right)\right\}\right]$$

$$\propto \exp\left[-\frac{1}{2}(B\mu^2 - 2\mu C + D)\right]$$

$$\propto \exp\left[-\frac{B}{2}\left(\mu - \frac{C}{B}\right)^2\right],$$

and $$\prod(\mu \mid \underset{\sim}{x}) = K \exp\left[-\frac{B}{2}\left(\mu - \frac{C}{B}\right)^2\right]$$

$$= \sqrt{\frac{B}{2\pi}} \exp\left[-\frac{B}{2}\left(\mu-\frac{C}{B}\right)^2\right].$$

Thus, the posterior distribution of μ is :

$$N\left(\frac{C}{B}, B^{-1}\right) \equiv N(\mu^*, \sigma^{2*})$$

where
$$B = \frac{n}{\sigma^2} + \frac{1}{\delta^2},$$

$$C = \frac{n\bar{x}}{\sigma^2} + \frac{m}{\delta^2},$$

$$D = \frac{n\bar{x}^2}{\sigma^2} + \frac{m^2}{\delta^2},$$

$$\mu^* = E(\mu \mid \underline{x}) = \frac{C}{B} = \frac{n\bar{x}\delta^2 + m\sigma^2}{n\delta^2 + \sigma^2}$$

$$= \frac{n\bar{x} + m\lambda}{n+\lambda} \qquad (2.3)$$

and,

$$\sigma^{2*} = B^{-1} = \frac{\delta^2 \sigma^2}{n\delta^2 + \sigma^2}$$

$$= \frac{\sigma^2}{n+\lambda}, \quad \lambda = \frac{\sigma^2}{\delta^2} \qquad (2.4)$$

Note that as $n \to \infty$ $\mu^* \to \bar{x}$, the maximum likelihood estimator (MLE) of μ which, of course, is what is expected. Further, as $\sigma^2 \to \infty$ which implies that as our prior information about μ 'becomes vaguer and vaguer' (Guttman, Wilks and Hunter, 1971), the posterior mean μ^* tends more and more to the sample mean $\bar{x}$ independent of the sample size.

Bayes risk for μ^*

From (1.12) we have Bayes risk

$$R^*(\mu^* \mid \mu) = \int_{-\infty}^{\infty} E_x\,[(\mu^* - \mu)^2],\ g(\mu)\, d\mu\,.$$

$$E(\mu^*) = \frac{n\mu + m\lambda}{n+\lambda}.$$

$$E_x[\mu^* - \mu)^2] = E_x\left[\mu^* - \frac{n\mu + m\lambda}{n+\lambda} + \frac{n\mu + m\lambda}{n+\lambda} - \mu\right]^2$$

$$= \text{Var}(\mu^*) + \frac{\lambda^2(\mu - m)^2}{(n+\lambda)^2}$$

$$= \frac{n\sigma^2 + \lambda^2(\mu - m)^2}{(n+\lambda)^2}.$$

$$R^*(\mu^*, \mu) = \frac{1}{(n+\lambda)^2} \int_{-\infty}^{\infty} [n\sigma^2 + \lambda^2(\mu - m)^2]\; g(\mu)\, d\mu$$

$$= \frac{n\sigma^2 + \lambda^2\delta^2}{(n+\lambda)^2}$$

$$= \frac{\sigma^2}{n+\lambda}.$$

(ii) σ known: state of 'in-ignorance' about μ

Suppose we are in a state of complete ignorance about μ and we represent our prior ignorance about μ by 'vague' prior (Jeffreys', 1961). In such a case, the likelihood (2.2) and the posterior distribution must be the same and hence we have

$$\prod(\mu \mid \underline{x}) = K \exp\left[-\frac{n}{2\sigma^2}(\bar{x} - \mu)^2\right],$$

where,
$$K^{-\cdot} = \int_{-\infty}^{\infty} \exp\left[-\frac{n}{2\sigma^2}(\bar{x} - \mu)^2\right] d\mu$$

$$= \sqrt{\frac{2\pi}{n}}\,\sigma.$$

Restoring the normalizing constant, we have

$$\prod(\mu \mid \underline{x}) = \frac{1}{\sqrt{2\pi}\;\sigma/\sqrt{n}} \exp\left[-\frac{1}{2\sigma^2/n}(\bar{x} - \mu)^2\right], -\infty < \mu \infty$$

which implies that the posterior distribution of μ is $N\left(\bar{x}, \dfrac{\sigma^2}{n}\right)$ and

$$\mu^* = E(\mu \mid \underline{x}) = \bar{x}.$$

(iii) μ, σ both unknown

Consider the family of improper or 'vague' priors

$$g(\mu, \sigma) \propto \frac{1}{\sigma^c}, \quad c > 0.$$

The joint prior distribution of μ and σ is given by

$$\prod(\mu, \sigma \mid \underline{x}) = \frac{K}{\sigma^{n+c}} \exp\left[-\frac{1}{2\sigma^2}\{A + n(\bar{x} - \mu)^2\}\right],$$

where
$$K^{-1} = \int_0^\infty \frac{\exp\left(-\frac{A}{2\sigma^2}\right)}{\sigma^{n+c}}$$

$$\left[\int_{-\infty}^{\infty} \exp\left\{-\frac{n}{2\sigma^2}(\bar{x} - \mu)^2\right\} d\mu\right] d\sigma$$

$$= \sqrt{\frac{2\pi}{n}} \int_0^\infty \frac{\exp\left(-\frac{A}{2\sigma^2}\right)}{\sigma^{n+c-1}} d\sigma$$

$$= \sqrt{\frac{\pi}{2n}} \int_0^\infty \frac{\exp\left(-\frac{A}{2\sigma^2}\right)}{(\sigma^2)^{(n+c-2)/2+1}} d\sigma^2$$

$$= \sqrt{\frac{\pi}{2n}}\, \Gamma\left(\frac{n+c-2}{2}\right)\left(\frac{2}{A}\right)^{(n+c-2)/2}$$

Restoring the constant K and integrating out σ from $\pi(\mu, \sigma \mid \underline{x})$, we have the marginal posterior of μ :

$$\prod_1(\mu \mid \underline{x}) = \sqrt{\frac{2n}{\pi}} \left(\frac{A}{2}\right)^{(n+c-2)/2} \left\{\Gamma\left(\frac{n+c-2}{2}\right)\right\}^{-1} \frac{1}{2}$$

$$\int_0^\infty \frac{\exp\left[-\frac{1}{2\sigma^2}\{A + n(\bar{x} - \mu)^2\}\right]}{(\sigma^2)^{(n+c-1)/2+1}} d\sigma^2$$

$$= \sqrt{\frac{2n}{\pi}} \left(\frac{A}{2}\right)^{(n+c-2)/2} \frac{\Gamma\left(\frac{n+c-1}{2}\right) 2^{(n+c-1)/2}}{2\,\Gamma\left(\frac{n+c-2}{2}\right) \{A + n(\bar{x}-\mu)^2\}^{(n+c-1)/2}}$$

$$= \sqrt{\frac{n}{A}} \frac{\Gamma\left(\frac{n+c-1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n+c-2}{2}\right)} \frac{1}{\left[1+\frac{n(\bar{x}-\mu)^2}{A}\right]^{(n+c-1)/2}}$$

$$= \sqrt{\frac{n}{A}} \frac{1}{B\left(\frac{1}{2}, \frac{n+c-2}{2}\right)\left\{1+\frac{n(\bar{x}-\mu)^2}{A}\right\}^{(n+c-1)/2}},$$

$$-\infty < \mu < \infty. \tag{2.5}$$

$$\mu^* = E(\mu \mid \underline{x})$$

$$= \sqrt{\frac{n}{A}} \cdot \frac{1}{B\left(\frac{1}{2}, \frac{n+c-2}{2}\right)} \int_{-\infty}^{\infty} \frac{\mu\, d\mu}{\left\{1+\frac{n(\bar{x}-\mu)^2}{A}\right\}^{(n+c-1)/2}}$$

Put $$\frac{\sqrt{n}\,(\bar{x}-\mu)}{\sqrt{A}} = \frac{t}{\sqrt{n+c-2}}$$

$$d\mu = -\sqrt{\frac{A}{n(n+c-2)}}\, dt,$$

and we have $$\mu^* = \frac{1}{B\left(\frac{1}{2}, \frac{n+c-2}{2}\right)\sqrt{n+c-2}}$$

$$\int_{-\infty}^{\infty} \frac{\left\{\bar{x} - \sqrt{\frac{A}{n(n+c-2)}}\, t\right\}}{\left(1+\frac{t^2}{n+c-2}\right)^{(n+c-2)/2+1}}\, dt$$

$$= \frac{\bar{x}}{B\left(\frac{1}{2}, \frac{n+c-2}{2}\right)\sqrt{n+c-2}} \int_{-\infty}^{\infty} \frac{dt}{\left(1+\frac{t^2}{n+c-2}\right)^{(n+c-2)/2+1}}$$

$$- \sqrt{\frac{A}{n\,(n+c-2)}}\; E(t)$$

$= \bar{x}$, since t is distributed as student's t with $(n+c-2)$ degrees of freedom. Thus, Bayes estimator of μ is the same as the MLE and UMVUE of μ and is independent of the choice of c.

Similarly, we obtain the marginal posterior of σ^2:

$$\prod (\sigma^2 \mid \underline{x}) = \int_{-\infty}^{\infty} \pi\,(\mu, \sigma \mid \underline{x})\, d\mu = K \sqrt{\frac{\pi}{2n}}\; \frac{\exp\left(-\frac{A}{2\sigma^2}\right)}{(\sigma^2)^{(n+c)/2}}$$

$$= \frac{\left(\frac{A}{2}\right)^{(n+c-2)/2}}{\Gamma\left(\frac{n+c-2}{2}\right)}\; \frac{\exp\left(-\frac{A}{2\sigma^2}\right)}{(\sigma^2)^{(n+c)/2}},\; 0 < \sigma < \infty \tag{2.6}$$

$$\sigma^{*^2} = E\,(\sigma^2 \mid \underline{x}) = \frac{\left(\frac{A}{2}\right)^{(n+c-2)/2}}{\Gamma\left(\frac{n+c-2}{2}\right)} \int_{0}^{\infty} \frac{\exp\left(\frac{-A}{2\sigma^2}\right)}{(\sigma^2)^{(n+c-4)/2+1}}\, d\sigma^2$$

$$= \frac{\Gamma\left(\frac{n+c-4}{2}\right)\left(\frac{A}{2}\right)^{(n+c-2)/2}}{\left(\frac{A}{2}\right)^{(n+c-4)/2}\;\Gamma\left(\frac{n+c-2}{2}\right)}$$

$$= \frac{A}{2}\;\frac{\Gamma\left(\frac{n+c-4}{2}\right)}{\Gamma\left(\frac{n+c-4}{2}+1\right)}$$

$$= \frac{A}{n+c-4}\,. \tag{2.7}$$

Thus, the Bayes estimator σ^2 identifies itself with the MLE for $c = 4$, with the UMVUE for $c = 3$ and for $c = 5$ (Hartigan prior, 1964), $\sigma*^2 = \frac{A}{n+1}$, the minimum mean-square of σ^2 (which we show in the following). Let KA be the min-mean-square estimator of σ^2, and let

$$M = E\{\,(KA - \sigma^2)^2\,\}$$

$$= K^2E(A^2) - 2K\sigma^2\,E(A) + \sigma^4$$

$$\frac{\partial N}{\partial K} = 0 \text{ yields } K = \frac{\sigma^2\,E(A)}{E(A^2)} = \frac{\sigma^2\,E(A)}{\mathrm{Var}\,(A) + \{E(A)\,]^2}\,;$$

It is well-known that $\frac{A}{\sigma^2} \sim \chi^2\ (n-1)$.

Hence, $\qquad E(A) = \sigma^2\,(n-1),\ \mathrm{Var}\,(A) = 2\sigma^4\,(n-1)$

and $\qquad K = \frac{n-1}{2(n-1)+(n-1)^2} = \frac{1}{n+1}\,.$

Thus, $\frac{A}{n+1}$ is the min-mean-squared-error estimator of σ^2.

Similarly, the marginal posterior of σ:

$$\prod\,(\sigma\,|\,\underline{x}) = \frac{A^{(n+c-2)/2}\,\exp\left(\frac{-A}{2\sigma^2}\right)}{2^{(n+c-4)/2}\,\Gamma\left(\frac{n+c-2}{2}\right)\sigma^{n+c-1}}\,,\ 0<\sigma<\infty.$$

$$\sigma^* = E\,(\sigma\,|\,\underline{x}) = \sqrt{\frac{A}{2}}\cdot\frac{\Gamma\left(\frac{n+c-3}{2}\right)}{\Gamma\left(\frac{n+c-2}{2}\right)}\,, \tag{2.8}$$

Example 2.1:

Fifteen items were subjected to a life testing experiment and the following failure times (hours) were recorded:

12.8	15.4	25.8	30.2	31.5
34.0	38.1	40.6	42.7	50.8
55.2	59.6	68.3	72.2	94.6

Suppose the failure time $X \sim N(\mu, \sigma = 20)$ and the prior knowledge about the product under test suggest that the mean failure time $\mu \sim N(m = 40, \delta = 10)$. We have $n = 15, \bar{x} = 44.79$. From (2.3) and (2.4),

$$\lambda = \frac{400}{100} = 4$$

$$\mu^* = \frac{15 \times 44.79 + 4 \times 40}{15 + 4} = 43.78,$$

$$\text{the posterior variance} = \sigma^{*^2} = \frac{1}{\dfrac{15}{500} + \dfrac{1}{100}} = 25.0.$$

If our prior knowledge is very vague/imprecise, we let $m \rightarrow \infty$, and $\mu^* = \bar{x} = 44.7$, and $\sigma^{*2} = \dfrac{400}{15} = 26.66$.

Now suppose that μ and σ are both unknown and we are in a state of in-ignorance about the prior distribution of (μ, σ) so that Jeffreys (1961) vague-prior $g(\mu, \sigma) \propto \dfrac{1}{\sigma}$ is applicable. Putting $c = 1$ in (2.7) and (2.8) we have

$$\sigma^{*^2} = \frac{A}{n-3} = \frac{6969.24}{12} = 580.77$$

$$\sigma^* = \sqrt{\frac{A}{2}} \frac{\Gamma\left(\dfrac{n-2}{2}\right)}{\Gamma\left(\dfrac{n-1}{2}\right)} = 23.61 \text{ and } \mu^* = \bar{x} = 44.79.$$

If we do not have the prior knowledge that σ is known and $\mu \sim N(m, \delta^2)$, the precision of σ^{*2} drops dramatically in this case from 0.0475 to 0.0017, the reason being that we have no prior information about (μ, σ) and we have agreed to use an 'ignorance' or 'vague' prior $g(\mu, \sigma) \propto \dfrac{1}{\sigma}$.

Robustness of the Posteriors

A random sample of size 30 observation was generated from a $N(\mu, \sigma^2)$ with $\mu = 20$ and $\sigma = 5$.

Sample

23.320	24.530	17.590	11.065	10.975
20.300	25.895	13.120	19.475	14.070
27.430	12.495	14.950	13.305	23.290
25.110	16.550	19.975	25.205	17.805
26.970	26.860	26.965	21.395	13.005
20.995	20.795	31.365	20.205	14.340

We have $\bar{x} = 19.945$, $A = 931.0165$, $n = 30$. Assuming (μ, σ) unknown and the joint prior $g(\mu, \sigma) \propto \frac{1}{\sigma^c}$, $c > 0$, the posteriors $\Pi(\mu \mid \underline{x})$ and $\Pi(\sigma \mid \underline{x})$ are plotted in Figs. 2.1 and 2.2, for $c = 0, 1, 2, 3$ for a wide range of values of μ and σ. The posteriors of μ are very much robust for variations in c while those of σ are less so. Robustness is also reflected in the Bayes estimators of μ and σ.

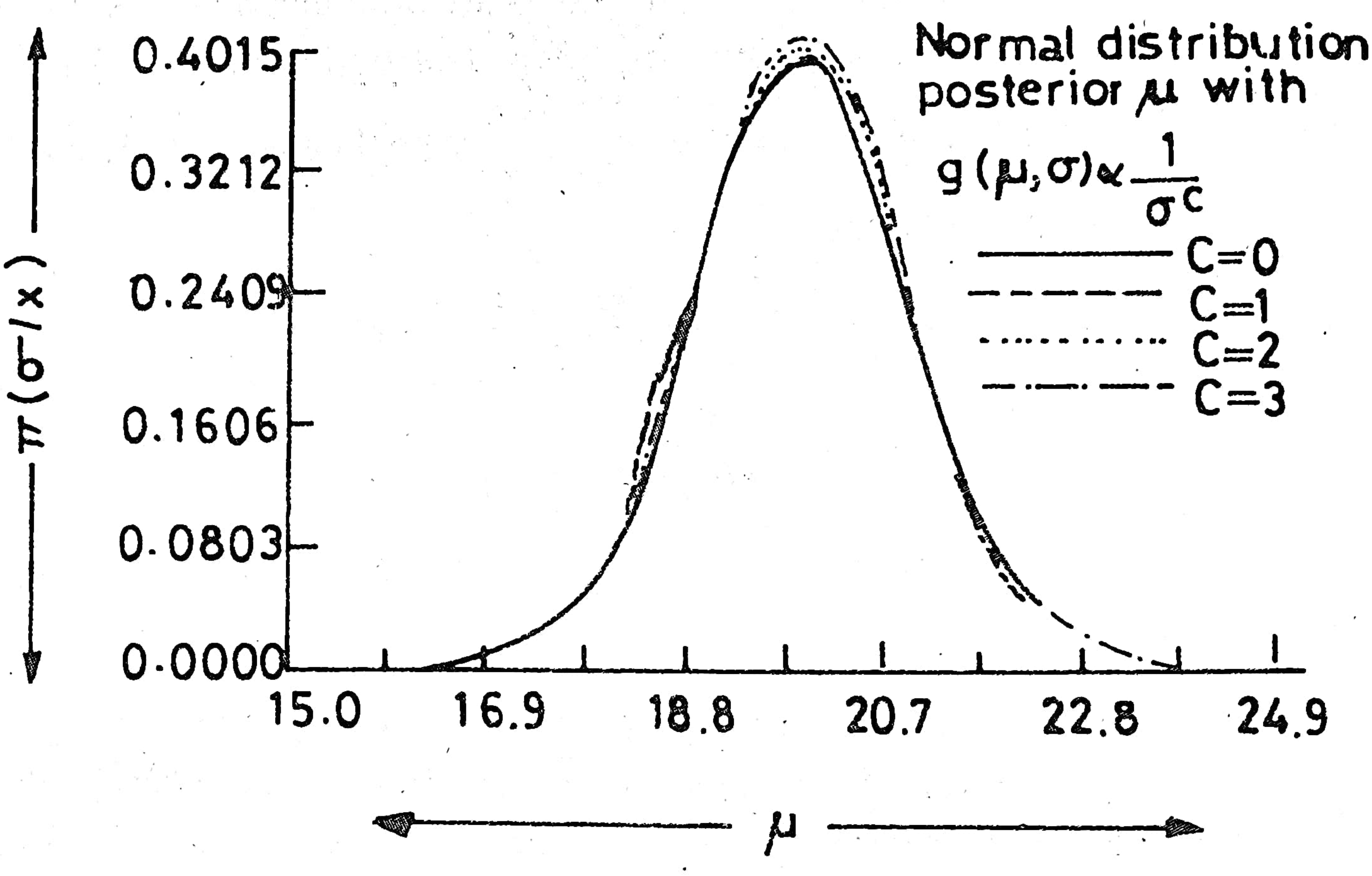


Fig. 2.1

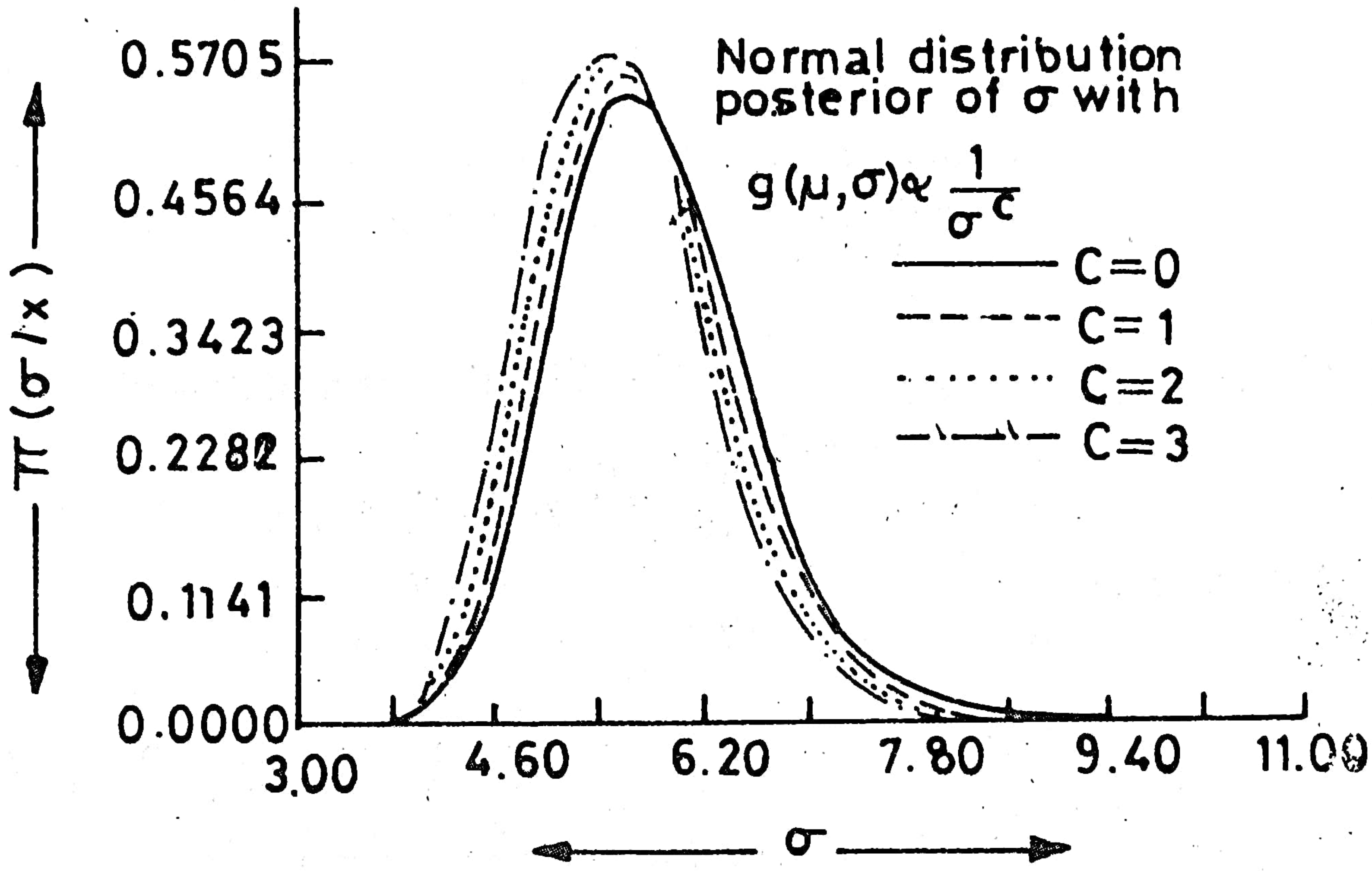


Fig. 2.2

Table 2.1 — Bayes Estimations of σ

c	σ	(Min/Max)
0	5.9268	
1	5.8180	0.9478
2	5.7151	
3	5.6174	

2.2 LOGNORMAL DISTRIBUTION

The lognormal distribution Sinha (1980) arises in different contexts such as in Physics (distributions of small particles), Economics (income distribution), Biology (growth of organisms), etc. A comprehensive treatment of lognormal distribution has been given by Aitchison and Brown (1957). The lognormal distribution serves as an appropriate model when the failure rate is rather high initially, and then decreases with time. The lognormal distribution arises when we assume that a positive random variable X is such that $Y = \log X$ is normally distributed with mean μ and variance σ^2.

Consider the lognormal pdf of a random variable X

$$f(x \mid \mu, \sigma) = \frac{x^{-1}}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{1}{2\sigma^2}(\log x - \mu)^2\right]$$

$$-\infty < \mu < \infty, \quad \sigma, x > 0. \tag{2.9}$$

Let (μ, σ) both be unknown.

Consider the likelihood function,

$$L(\mu, \sigma \mid \underset{\sim}{x}) \propto \frac{1}{\sigma^n} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(\log x_i - \mu)^2\right] \tag{2.10}$$

and the non-informative joint prior,

$$g(\mu, \sigma) \propto \frac{1}{\sigma}.$$

Comparing (2.2) and (2.10) it follows that we may obtain the estimators (μ^*, σ^*) by replacing x_i by $\log x_i$ and $c = 1$ in case (iii) of Section 2.1.

We will have,

$$\mu^* = \frac{\sum_{i=1}^{n} \log x_i}{n},$$

and
$$\sigma^* = \sqrt{\frac{\lambda}{2}}\left[\frac{\Gamma\left(\frac{n-2}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}\right], \quad n > 2$$

where
$$\lambda = \sum_{i=1}^{n}(\log x_i)^2 - \frac{\left(\sum_i \log x_i\right)^2}{n}$$

2.3 MULTINORMAL DISTRIBUTION

The multinormal distribution is a generalization of univariate normal distribution to the case of n-variates ($n \geq 2$). If the i^{th} variate x_i has mean

μ_i and covariance matrix (m_{ij}), $i, j = 1, 2 \cdots, n$ with an inverse (M_{ij}), the multinormal distribution has the frequency function,

$$f(x_1, x_2, \cdots, x_n \mid \mu) = (2\pi)^{-n/2} \, |M|^{-1/2} \exp\left[-\frac{1}{2}(x-\mu)' M^{-1} (x-\mu)\right] \quad (2.11)$$

where $\quad x' = (x_1, x_2, \cdots, x_n)$, $\mu' = (\mu_1, \mu_2, \cdots, \mu_n)$

and M is the $(n \times n)$ covariance matrix (m_{ij}).

Let $$g(\mu) \propto \exp\left[-\frac{1}{2}(\mu - a)' K^{-1} (\mu - a)\right] \quad (2.12)$$

be the prior distribution of μ.

Combining (2.11) and (2.12), the posterior distribution of μ is given by

$$\Pi(\mu \mid x) \propto \exp\left[-\frac{1}{2}\{(x-\mu)' M^{-1}(x-\mu) + (\mu - a)' K^{-1}(\mu - a)\}\right].$$

Consider the expression

$$\begin{aligned}
&\{(x-\mu)' M^{-1}(x-\mu) + (\mu - a)' K^{-1}(\mu - a)\} \\
&= x'M^{-1}x - x'M^{-1}\mu - \mu'M^{-1}x + \mu'M^{-1}\mu \\
&\quad + \mu'K^{-1}\mu - \mu'K^{-1}a - a'K^{-1}\mu + a'K^{-1}a \\
&= x'M^{1}x - xM^{-1}\mu - \mu'M^{-1}x + \mu'M^{-1}\mu \\
&\quad + \mu'K^{-1}\mu - \mu'K^{-1}a - a'K^{-1}\mu + a'K^{-1}a \\
&= x'M^{-1}x - \mu'(K^{-1}a + M^{-1}x) - (a'K^{-1} + x'M^{-1})\mu \\
&\quad + \mu'(M^{-1} + K^{-1})\mu + a'K^{-1}a \\
&= x'M^{-1}x + a'K^{-1}a - \theta'G^{-1}\theta + (\mu - \theta)' G^{-1}(\mu - \theta)
\end{aligned}$$

where, $$M^{-1} + K^{-1} = G^{-1} \quad (2.13)$$

$$K^{-1}a + M^{-1}x = G^{-1}\theta \quad (2.14)$$

$$a'K^{-1} + x'M^{-1} = \theta'G^{-1}.$$

Absorbing $x'M^{-1}x + a'K^{-1}a - \theta'G^{-1}\theta)$ in the constant, we have

$$\Pi(\mu \mid x) \propto \exp\left[-\frac{1}{2}(\mu - \theta)' G^{-1}(\mu - \theta)\right]$$

and Bayes estimator,

$$\mu^* = \theta.$$

From (2.13) and (2.14),

$$\begin{aligned}
\theta &= G\,(K^{-1}a + M^{-1}x) \\
&= (M^{-1}+K^{-1})^{-1}\;(K^{-1}a+M^{-1}x) \\
&= x-(M^{-1}+K^{-1})^{-1}\,(M^{-1}+K^{-1})x \\
&\quad + (M^{-1}+K^{-1})^{-1}\cdot K^{-1}a + (M^{-1}+K^{-1})^{-1}\cdot M^{-1}x \\
&= x-(M^{-1}+K^{-1})^{-1}\,M^{-1}x-(M^{-1}+K^{-1})^{-1}\cdot K^{-1}x \\
&\quad + (M^{-1}+K^{-1})^{-1}\cdot K^{-1}a + (M^{-1}+K^{-1})^{-1}\cdot M^{-1}x \\
&= x+(M^{-1}+K^{-1})^{-1}\,K^{-1}\,(a-x) \\
&= x+\left[K\,(M^{-1}+K^{-1})\right]^{-1}(a-x) \\
&= x+(I+KM^{-1})^{-1}\,(a-x) \qquad (2.15)
\end{aligned}$$

(Kimeldorf–Jones, 1967)

Example 2.2:

Let μ_1, μ_2, μ_3 be independently normally distributed with prior means $a' = (2, 4, 7)$, $x' = (1, 3, 8]$,

$$K = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{bmatrix} \text{ and } M = \begin{bmatrix} \frac{1}{4} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{4}{5} \end{bmatrix}.$$

Find the posterior mean vector,

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix}.$$

We have,

$$I + KM^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} 4 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & \frac{5}{4} \end{bmatrix}$$

$$= \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 6 \end{bmatrix}$$

$$a - x = \begin{bmatrix} 2 \\ 4 \\ 7 \end{bmatrix} - \begin{bmatrix} 1 \\ 3 \\ 8 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$$

From (2.15)

$$\theta = \begin{bmatrix} 1 \\ 3 \\ 8 \end{bmatrix} + \begin{bmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{6} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{4}{3} \\ \frac{10}{3} \\ \frac{47}{6} \end{bmatrix} = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix}$$

2.4 BINOMIAL DISTRIBUTION

Suppose a market research firm wants to estimate the proportion p of families in an area that uses a certain brand of detergent. A random sample of n families were interviewed and let X be the number of families who use the product. The probability distribution of X is given by

$$f(x \mid p) = \binom{n}{x} p_x q^{n-x}, \quad q = 1 - p; \quad x = 0, 1, 2, \cdots, n. \qquad (2.16)$$

Suppose the interview is extended over a period of N weeks. Let $(x_1, x_2, \cdots, x_N) = \underline{x}$ be a sample of N independent observations from this distribution. The likelihood function

$$L(x_i \mid p) = \prod_{i=1}^{N} \binom{n}{x_i} p^{x_i} q^{n-x_i}$$

$$= p^S \ (1-p)^{Nn-S} \prod_{i=1}^{N} \binom{n}{x_i},$$

$$S = \sum_{i=1}^{N} x_i \, .$$

Consider the conjugate prior distribution of p:

$$g\,(p) \propto p^{a-1} \ (1-p)^{b-1} \quad a,\ b > 0. \tag{2.17}$$

The posterior distribution of p is given by

$$\prod \ (p \mid x) = Kp^{S+a-1} \ (1-p)^{Nn+b-S-1}$$

where $$K^{-1} = \int_0^1 \prod \ (p \mid x) \ dp = B\,(S+a,\ Nn+b-S).$$

Restoring the normalizing constant, the posterior distribution of p is given by

$$\Pi \ (p \mid x) = \frac{1}{B\,(S+a,\ Nn+b-S)} \ p^{S+a-1} \ (1-p)^{Nn+b-S+1} \quad 0<p<1.$$

Hence, the Bayes estimator

$$p* = E\,(p \mid \underline{x}) = \frac{B\,(S+a-1,\ Nn+b-S)}{B\,(S+a,\ Nn+b-S)}$$

$$= \frac{S+a}{Nn+a+b} \tag{2.18}$$

The parameters $(a,\ b)$ are usually unknown. One may obtain the marginal distribution of X and estimate $(a,\ b)$ by the method of moments. (2.17) represents a Beta-prior. Restoring the normalizing constant and combining with (2.16), the joint distribution $(X,\ a,\ b)$ is given by

$$h\,(x \mid a,\ b) = \frac{\binom{n}{x}}{B\,(a,\ b)} \ B\,(x+a,\ n-x+b)$$

$$= K \binom{x+a-1}{a-1} \binom{n-x+b-1}{b-1} \tag{2.19}$$

where, $$K^{-1} = \binom{n+a+b-1}{n+b-1}$$

The probability distribution (2.19) is known as a Beta-binomial distribution, $B(x, n, a, b)$. From (2.19) it follows that

$$\binom{n+a+b-1}{a+b-1} = \sum_{x=0}^{n} \binom{x+a-1}{a-1} \binom{n-x+b-1}{b-1}$$

$$= \sum_{x=0}^{n} \frac{(a+x-1)!\ (n-x+b-1)!}{(a-1)!\ (b-1)!\ x!\ (n-x)!}. \qquad (2.20)$$

$$E(X) = \sum_{x=0}^{n} xh(x \mid a, b)$$

$$= K \sum_{x=1}^{n} \frac{(a+x-1)!\ (b+n-x-1)!}{(a-1)!\ (b-1)!\ (x-1)!\ (n-x)!}$$

$$= Ka \sum_{y=0}^{n-1} \frac{(a+y)!\ (b+n-y-2)!}{a!\ (b-1)!\ y!\ (n-y-1)!}$$

$$= Ka \sum_{y=0}^{m} \frac{(a+1+y-1)!\ (b+m-y-1)!}{(a+1-1)!\ (b-1)!\ (m-y)!\ y!}$$

and comparing (2.19) we have

$$E(X) = Ka \binom{m+a+1+b-1}{a+1\,b-1}$$

$$= Ka \binom{n+a+b-1}{a+b} = \frac{na}{a+b}. \qquad (2.21)$$

Equating (2.21) to the sample mean $\bar{x}$ we have the estimating equation

$$\bar{x} = \frac{na}{a+b}.$$

Substituting in (2.18)

$$p^* = \frac{N\bar{x}+a}{Nn + \dfrac{na}{\bar{x}}}$$

$$= \frac{\bar{x}\left(N + \frac{a}{\bar{x}}\right)}{n\left(N + \frac{a}{\bar{x}}\right)}$$

$$= \frac{\bar{x}}{n}$$

= the maximum likelihood (also the moment) estimator of p.

Thus, we observe an interesting result that independent of the parameters (a, b) of the Beta-prior (2.17), Bayes estimator of p identifies itself with the maximum likelihood as well as the moment estimator of p.

The corresponding estimators under non-informative prior ($a = b = 1$), Jeffreys' $\left(1961: a = b = \frac{1}{2}\right)$ or ALI prior ($a = b = 0$; Hartigan, 1964) may be obtained by substituting suitable values for (a, b) in (2.18).

2.5 MULTINOMIAL DISTRIBUTION

Multinomial distribution is a generalization of the binomial distribution in which each trial permits more than just two outcomes. Suppose a sequence of n independent trials results in k mutually exclusive events E_1, E_2, E_k and let $p_i = P(E_i)$, and we are interested in the probability that the event E_i occurs exactly x_i times, $i = 1, 2, \cdots, k$.

The joint probability distribution of the random variables $(X_1, X_2, \cdots, X_K)$ is known as the multinomial distribution, and is given by

$$P(x_1, x_2, \cdots, x_k \mid p_1, p_2, \cdots, p_k) = \frac{n!\, p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}}{x_1!\, x_2! \cdots x_k!},$$

$$\sum_{i=1}^{k} p_i = 1, \quad \sum_{i=1}^{k} x_i = n.$$

Consider the natural conjugate family of prior for $(p_1, p_2, \cdots, p_k)$ given by

$$g(p_1, p_2, \cdots, p_k) \propto p_1^{\theta_1 - 1}\, p_2^{\theta_2 - 1}, \cdots p_k^{\theta_k - 1}, \quad \theta_i > 0.$$

Combining the prior and the probability distribution of $(x_1, x_2, \cdots x_k)$ we have the joint posterior distribution,

$$\Pi\ (p_1, p_2, \cdots, p_k \mid x_1, x_2, \cdots, x_k) \propto p_1^{x_1+\theta_1-1}\ p_2^{x_2+\theta_2-1} \ \ldots\ p_k^{x_k+\theta_k+1}$$

which belongs to the same family as $g\ (p_1, p_2, \cdots, p_k)$. We have shown in (1.10) that the prior expectation,

$$E\ (p_i) = \frac{\theta_i}{\theta},\ \theta = \sum_{i=1}^{k} \theta_i\,.$$

Replacing θ_i by $\theta_i + x_i$, we have Bayes estimator

$$p_i^* = E\ (p_i \mid x_1, x_2, \cdots, x_k) = \frac{\theta_i + x_i}{\theta + n}\,.$$

Note that on putting $k = 2$, $p_1 = p$, $p_2 = 1 - p_1 = q$, we have the well-known binomial distribution

$$p(x) = \binom{n}{x} p^x\ (1-p)^{n-x},\quad x = 0, 1, \cdots, n;\quad 0 < p < 1,$$

the beta-prior,

$$g(p) \propto p_i^{\theta-1}\ (1-p)^{\theta_2-1}$$

and Bayes estimate of p is given by

$$p^* = \frac{\theta_1 + x}{n + \theta_1 + \theta_2}\,.$$

Multinomial distribution has useful applications in Actuarial Science (London, 1985).

2.6 POISSON DISTRIBUTION

Let X be the number of suicides committed in a large community over a period of time. Since suicide is a rare event, we may assume that the frequency of suicides follows the Poisson probability distribution,

$$f(x \mid \lambda) = \exp\ (-\lambda)\ \frac{\lambda^x}{x!}\quad x = 0, 1, 2, \cdots. \qquad (2.22)$$

Different members of the population may have different degrees of suicide-proneness represented by different values of λ in the Poisson distribution.

Consider the NCP for λ given by

$$g(\lambda \mid a, b) = \frac{b^a}{\Gamma(a)} \lambda^{a-1} \exp(-b\lambda), \quad a, b, \lambda > 0. \tag{2.23}$$

Let $x = (x_1, x_2, \cdots, x_n)$ be a random sample from the Poisson distribution (2.22). The likelihood function,

$$L(\lambda \mid \underline{x}) = \frac{\exp(-n\lambda)\, \lambda^S}{x_1!\, x_2! \cdots x_n!}, \quad S = \sum_{i=1}^{n} x_i .$$

The posterior distribution

$$\prod(\lambda \mid \underline{x}) = K \exp[-\lambda\, (n+b)]\, \lambda^{S+a+1}$$

$$= \frac{(n+b)^{S+a}}{\Gamma(S+a)} \exp[-\lambda\,(n+b)]\, \lambda^{S+a-1}$$

$$\lambda^* = E(\lambda \mid \underline{x}) = \frac{(n+b)^{S+a}}{\Gamma(S+a)} \frac{\Gamma(S+a+1)}{(n+b)^{S+a+1}}$$

$$= \frac{S+a}{n+b} \tag{2.24}$$

Example 2.3:

Let $\underline{x} = (x_1, x_2, \cdots, x_{15})$ be a random sample from the Poisson pdf

$$f(x \mid \lambda) = \frac{\exp(-\lambda)\, \lambda^x}{x!}, \quad x = 0, 1, 2, \cdots$$

and let the prior distribution of λ be

$$g(\lambda) = 4\lambda^2 \exp(-2\lambda), \; \lambda > 0.$$

Obtain Bayes risk for the corresponding estimator of λ given

$$S = \sum_{i=1}^{15} x_i = 235.$$

We have $n = 15$, $a = 3$, $b = 2$. From (2.24),

$$\lambda^* = \frac{235+3}{15+2} = 14 = \frac{S+a}{n+b}$$

$$E_x(\lambda^*) = \frac{n\lambda + a}{n+b} = \frac{15\lambda + 3}{17}$$

$$\text{Var}_X(\lambda^*) = \frac{n\lambda}{(n+b)^2} = \frac{15\lambda}{289}$$

$$E_X[(\lambda^* - \lambda)^2] = \text{Var}(\lambda^*) + \left[\frac{15\lambda+3}{17} - \lambda\right]^2$$

$$= \frac{15\lambda + (3-2\lambda)^2}{289}$$

$$= \frac{4\lambda^2 + 3\lambda + 9}{289}.$$

From (1.12) we have Bayes risk

$$R^*(\lambda^*, \lambda) = \frac{1}{289}\,[4\,E(\lambda)^2 + 3\,E(\lambda) + 9]$$

$$E(\lambda) = \int_0^\infty g(\lambda)\lambda\, d\lambda = \frac{3}{2}.$$

Similarly, $E(\lambda^2) = 3$ and

$$R^*(\lambda^*, \lambda) = \frac{1}{289}\left[12 + \frac{9}{2} + 9\right] = \frac{3}{34}.$$

2.7 EXPONENTIAL DISTRIBUTION

(i) One-parameter exponential:

The pdf of the random variable X is given by

$$f(x \mid \theta) = \frac{1}{\theta}\exp\left(-\frac{x}{\theta}\right), \quad x,\ \theta > 0 \tag{2.25}$$

and the likelihood function,

$$L(\underline{x} \mid \theta) = \frac{1}{\theta^n}\exp\left(-\frac{S}{\theta}\right), \quad S = \sum_{i=1}^{n} x_i.$$

Consider the inverted gamma prior,

$$g(\theta \mid a, b) = \frac{a^b}{\Gamma(b)} \frac{\exp\left(-\frac{a}{\theta}\right)}{\theta^{b+1}}, \quad a, b, \theta > 0 \text{ (Raiffa and Schlaifer, 1961)}$$

which is conjugate to the exponential family (2.25).

The posterior distribution,

$$\prod(\theta \mid \underline{x}) = \frac{(S+a)^{n+b} \exp\left(-\frac{S+a}{\theta}\right)}{\Gamma(n+b)\ \theta^{n+b+1}}$$

and Bayes estimator,

$$\theta^* = E(\theta \mid \underline{x})$$

$$= \frac{(S+a)^{n+b}\ \Gamma(n+b-1)}{\Gamma(n+b)\ (S+a)^{n+b-1}}$$

$$= \frac{S+a}{n+b-1}.$$

Using Jeffreys (1961) prior, $g(\theta) \propto \frac{1}{\theta}$, the posterior distribution

$$\prod(\theta \mid \underline{x}) = \frac{K}{\theta^{n+1}} \exp\left(-\frac{S}{\theta}\right)$$

$$= \frac{S^n \exp\left(-\frac{S}{\theta}\right)}{\Gamma(n)\ \theta^{n+1}} \tag{2.26}$$

$\theta^* = E(\theta \mid \underline{x}) = \dfrac{S}{n-1}$ and as $n \to \infty$,

$\theta^* \to \bar{x}$, the MLE/UMVUE of θ,

$$E(\theta^2 \mid \underline{x}) = \frac{S^n\, \Gamma(n-2)}{\Gamma(n)\, S^{n-2}} = \frac{S^2}{(n-1)\ (n-2)}.$$

posterior variance

$$V_1(\theta \mid \underline{x}) = \frac{S^2}{(n-1)\ (n-2)} - \frac{S^2}{(n-1)^2}$$

$$= \frac{S^2}{(n-1)^2\ (n-2)}.$$

Consider ALI prior (Hartigan, 1964)

$$g(\theta) \propto \frac{1}{\theta^2}.$$

Proceeding as above, we obtain

$\theta^* = \frac{S}{n} = \bar{x}$, the MLE/UMVUE of θ, irrespective of the sample size n

and the posterior variance,

$$V_2(\theta \mid \underset{\sim}{x}) = \frac{S^2}{n^2(n-1)}$$

$$\frac{V_2(\theta \mid \underset{\sim}{x})}{V_1(\theta \mid \underset{\sim}{x})} = \frac{(n-1)\ (n-2)}{n^2}$$

$$= \left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right) < 1 \text{ for } n > 2.$$

Thus, the posterior variance $V(\theta \mid \underset{\sim}{x})$ under ALI prior (Hartigan, 1964) is less than its counterpart under Jeffreys (1961) invariance prior.

(ii) Two-parameter exponential

We now consider the two-parameter exponential density function

$$f(x \mid \mu, \theta) = \frac{1}{\theta} \exp\left(-\frac{x-\mu}{\theta}\right), \quad -\infty < \mu < x < \infty, \quad \theta > 0 \tag{2.27}$$

and the vague-prior,

$$g(\mu, \theta) \propto \frac{1}{\theta}.$$

The likelihood function,

$$L(\mu, \theta \mid \underset{\sim}{x}) = \frac{1}{\theta^n} \exp\left[-\frac{1}{\theta}\sum_{i=1}^{n}\{x_i - \mu\}\right]$$

$$= \frac{1}{\theta^n} \exp\left[-\frac{1}{\theta}\{S + n(x_{(1)} - \mu)\}\right]$$

where $x_{(1)}$ is the first order statistic in the sample $\underset{\sim}{x} = (x_1, x_2, \cdots, x_n)$ and

$$S = \sum_{i=1}^{n} \{x_i - x_{(1)}\}.$$

The joint posterior distribution,

$$\prod (\mu, \theta \mid \underset{\sim}{x}) \propto L(\mu, \theta \mid \underset{\sim}{x})\; g(\theta)$$

$$\propto \frac{1}{\theta^{n+1}} \exp\left[-\frac{1}{\theta}\{S + n(x_{(1)} - \mu)\}\right]. \tag{2.28}$$

The marginal posterior of μ is given by

$$\prod_1 (\mu \mid \underset{\sim}{x}) = \int_0^\infty \prod (\mu, \theta \mid \underset{\sim}{x})\, d\theta$$

$$\propto \int_0^\infty \frac{\exp\left[-\frac{1}{\theta}\{S + n\,(x_{(1)} - \mu)\}\right]}{\theta^{n+1}}\, d\theta$$

$$= \frac{K}{\left[S + n(x_{(1)} - \mu)\right]^n} \qquad -\infty < \mu < x_{(1)} \tag{2.29}$$

where $$K^{-1} = \int_{-\infty}^{x_{(1)}} \frac{d\mu}{\left[S + n(x_{(1)} - \mu)\right]^n}.$$

Let $S + n\,(x_{(1)} - \mu) = \nu$. Then

$$d\mu = -\frac{d\nu}{n}$$

$$K^{-1} = -\frac{1}{n}\int_\infty^S \frac{d\nu}{\nu^n} = \frac{1}{n\,(n-1)}\left[\frac{1}{\nu^{n-1}}\right]_\infty^S$$

$$= \frac{1}{n\,(n-1)\; S^{n-1}}.$$

Substituting K in (2.29), the marginal posterior

$$\prod_1 (\mu \mid \underset{\sim}{x}) = \frac{n\,(n-1)\,S^{n-1}}{\left[S + n\,(x_{(1)} - \mu)\right]^n} \qquad -\infty < \mu < x_{(1)}.$$

$$\mu^* = \int_{-\infty}^{x_{(1)}} \prod_1 (\mu \mid \underset{\sim}{x})\, d\mu$$

$$= (n-1)\, S^{n-1} \int_{-\infty}^{x_{(1)}} \frac{n\mu\, d\mu}{\left[S + n\,(x_{(1)} - \mu)\right]^n}$$

$$= -\frac{(n-1)\, S^{n-1}}{n} \int_{\infty}^{S} \frac{\{S + nx_{(1)} - \nu\}}{\nu^n}\, d\nu$$

$$= \frac{(n-1)\, S^{n-1}}{n} \left[\frac{S + n\, x_{(1)}}{(n-1)\, S^{n-1}} - \frac{1}{(n-2)\, S^{n-2}}\right]$$

$$= \frac{n-1}{n} \left[\frac{S + n\, x_{(1)}}{n-1} - \frac{S}{n-2}\right]$$

$$= \frac{1}{n\,(n-2)} \left[n\,(n-2)\, x_{(1)} - S\right]$$

$$= x_{(1)} - \frac{S}{n\,(n-2)}. \qquad (2.30)$$

$\prod_2 (\theta \mid \underset{\sim}{x})$ = marginal posterior of θ

$$\propto \int_{-\infty}^{x_{(1)}} \prod (\mu,\ \theta \mid \underset{\sim}{x})\, d\mu$$

$$\propto \frac{1}{\theta^{n+1}} \int_{-\infty}^{x_{(1)}} \exp\left[-\frac{1}{\theta}\, \{S + n\,(x_{(1)} - \mu)\}\right] d\mu$$

$$\propto \frac{1}{\theta^{n+1}} \int_{S}^{\infty} \exp\left(-\frac{\nu}{\theta}\right) d\nu$$

$$= \frac{C}{\theta^n} \exp\left(-\frac{S}{\theta}\right) \qquad (2.30)$$

$$= \frac{S^{n-1} \exp\left(-\dfrac{S}{\theta}\right)}{\Gamma\,(n-1)\, \theta^n} \qquad \theta > 0, \quad n > 1.$$

$$\theta^* = \frac{S^{n-1}}{\Gamma(n-1)} \int_0^\infty \frac{\exp\left(-\frac{S}{\theta}\right)}{\theta^{n-1}}\, d\theta$$

$$= \frac{S\Gamma(n-2)}{\Gamma(n-1)} = \frac{S}{n-2}, \quad n > 2. \tag{2.31}$$

(iii) Left truncated exponential distribution

In the context of life testing experiments, the model (2.27) is unrealistic. We now consider the more realistic model

$$f(x \mid \mu, \theta) = \frac{1}{\theta} \exp\left[-\frac{1}{\theta}(x-\mu)\right], \quad 0 < \mu < x \le \infty,\ \theta > 0 \tag{2.32}$$

and the same prior

$$g(\mu, \theta) \propto \frac{1}{\theta}.$$

The normalizing constant K in (2.29) will change to

$$K_1^{-1} = \int_0^{x_{(1)}} \frac{d\mu}{\left[S + n(x_{(1)} - \mu)\right]^n}$$

$$= -\frac{1}{n} \int_{S+nx_{(1)}}^{S} \frac{dv}{v^n}$$

$$= \frac{1}{n(n-1)} \left[\frac{1}{S^{n-1}} - \frac{1}{(S+nx_{(1)})^{n-1}}\right]$$

$$= \frac{1}{n(n-1)S^{n-1}} \left[1 - \left\{\frac{S}{S+nx_{(1)}}\right\}^{n-1}\right]$$

$$\prod\nolimits_1 (\mu \mid \underline{x}) = \frac{K_1}{\left[S + n(x_{(1)} - \mu)\right]^n} \quad 0 < \mu < x_{(1)} \tag{2.33}$$

where $$K_1 = n(n-1)S^{n-1} \left[1 - \left\{\frac{S}{S+nx_{(1)}}\right\}^{n-1}\right]^{-1}$$

$$\mu^* = \frac{K_1}{n} \int_0^{x_{(1)}} \frac{n\mu \, d\mu}{\left[S + n\left(x_{(1)} - \mu\right)\right]^n}$$

$$= \frac{K_1}{n^2} \left[- \int_{S+nx_{(1)}}^{S} \frac{\left\{S + nx_{(1)} - \nu\right\}}{\nu^n} \, d\nu \right]$$

$$= \frac{K_1}{n^2} \left[-\left\{S + nx_{(1)}\right\} \int_{S+nx_{(1)}}^{S} \frac{d\nu}{\nu^n} + \int_{S+nx_{(1)}}^{S} \frac{d\nu}{\nu^{n-1}} \right]$$

$$= \frac{K_1}{n^2} \left[\frac{\left\{S + nx_{(1)}\right\}}{n-1} \left\{ \frac{1}{S^{n-1}} - \frac{1}{\left(S + nx_{(1)}\right)^{n-1}} \right\} - \frac{1}{n-2} \left\{ \frac{1}{S^{n-2}} - \frac{1}{\left(S + nx_{(1)}\right)^{n-2}} \right\} \right]$$

$$= \frac{K_1}{n^2} \left[\frac{\left\{S + nx_{(1)}\right\}}{(n-1)\, S^{n-1}} \left\{ 1 - \left(\frac{S}{S + nx_{(1)}} \right)^{n-1} \right\} - \frac{1}{(n-2)\, S^{n-2}} \left\{ 1 - \left(\frac{S}{S + nx_{(1)}} \right)^{n-2} \right\} \right]$$

$$= \frac{S + nx_{(1)}}{n} - \frac{K_1}{n^2 (n-2)\, S^{n-2}} \left\{ 1 - \left(\frac{S}{S + nx_{(1)}} \right)^{n-2} \right\}.$$

Substituting K_1, we obtain

$$\mu^* = x_{(1)} + \frac{S}{n} \left[1 - \frac{n-1}{n-2} \left\{ \frac{1 - \left(\dfrac{S}{S + nx_{(1)}} \right)^{n-2}}{1 - \left(\dfrac{S}{S + nx_{(1)}} \right)^{n-1}} \right\} \right]. \qquad (2.34)$$

Similarly,

$$\Pi_2(\theta \mid \underline{x}) \propto \frac{1}{\theta^{n+1}} \int_0^{x_{(1)}} \exp\left[-\frac{1}{\theta}\{S + n(x_{(1)} - \mu)\}\right] d\mu$$

$$\propto \frac{1}{\theta^{n+1}} \int_S^{S+nx_{(1)}} \exp\left(-\frac{\nu}{\theta}\right) d\theta$$

$$= \frac{C_1}{\theta^n}\left[\exp\left(-\frac{S}{\theta}\right) - \exp\left\{-\frac{S+nx_{(1)}}{\theta}\right\}\right] \tag{2.35}$$

where, $$C_1^{-1} = \int_0^\infty \frac{\exp\left(-\frac{S}{\theta}\right) - \exp\left\{-\frac{S+nx_{(1)}}{\theta}\right\}}{\theta^n} d\theta$$

$$= \Gamma(n-1)\left[\frac{1}{S^{n-1}} - \frac{1}{\{S+nx_{(1)}\}^{n-1}}\right]$$

$$= \frac{\Gamma(n-1)}{S^{n-1}}\left[1 - \left\{\frac{S}{S+nx_{(1)}}\right\}^{n-1}\right]$$

$$= \frac{\Gamma(n+1)}{K_1}$$

Substituting $C_1 = \dfrac{K_1}{\Gamma(n+1)}$ in (2.35), we get

$$\Pi_2(\theta \mid \underline{x}) = \frac{K_1}{\Gamma(n+1)}\left[\frac{\exp\left(\frac{S}{\theta}\right) - \exp\left\{-\frac{S+nx_{(1)}}{\theta}\right\}}{\theta^n}\right] \qquad \theta > 0.$$

$$\theta^* = \frac{K_1}{\Gamma(n+1)}\left[\int_0^\infty \frac{\exp\left(\frac{S}{\theta}\right)d\theta}{\theta^{n-1}} - \int_0^\infty \frac{\exp\left\{-\frac{S+nx_{(1)}}{\theta}\right\}}{\theta^{n-1}}d\theta\right]$$

$$= \frac{K_1\,\Gamma(n-2)}{\Gamma(n+1)}\left[\frac{1}{S^{n-2}} - \frac{1}{\left\{S+nx_{(1)}\right\}^{n-2}}\right]$$

$$= \frac{S}{n-2}\left[\frac{1-\left\{\dfrac{S}{S+nx_{(1)}}\right\}^{n-2}}{1-\left\{\dfrac{S}{S+nx_{(1)}}\right\}^{n-1}}\right] \tag{2.36}$$

2.8 WEIBULL DISTRIBUTION

The literature abounds with models assumed suitable for life testing and reliability situation. The simplest and the most widely used failure time distribution is the exponential distribution (2.25) and (2.27). An interesting property of the exponential distribution is that the distribution is 'forgetful' or 'has no memory'. What it means, of course, is that if a unit has survived T number of hours, then the probability of the unit surviving an additional h hours is the same as the probability of a new item surviving h hours. However, not all items have this unusual property that 'they do not age'. There are several situations where the failure rate may increase (or decrease) with age.

Weibull distribution is one such distribution which takes 'fatigue' or 'age' factor into account and has been widely used in life testing and reliability problems. This distribution has been named after the Swedish Scientist Weibull (1939), who first proposed it in connection with his studies on strength of materials.

The probability density function (pdf) of Weibull random variable X is given by

$$f(x\,|\,\theta, p) = \frac{p}{\theta}\, x^{p-1} \exp(-x^p/\theta), \quad x, p, \theta > 0. \tag{2.37}$$

Note that for $p = 1$, we have the exponential pdf (2.25). p is called the shape parameter.

Let $\underline{x} = (x_1, x_2, \cdots, x_n)$ be a sample of n independent observations on X.

The likelihood function

$$L(\theta \mid \underline{x}) \propto p^n \theta^{-n} \lambda^{p-1} \exp\left(-\sum_{i=1}^{n} x_i^p / \theta\right), \quad \lambda = \prod_{i=1}^{n} (x_i). \tag{2.38}$$

Consider Jeffreys (1961) prior $h(\theta, p) \propto \dfrac{1}{\theta p}$. (2.39)

From (2.38) and (2.39) we obtain the joint posterior of (θ, p)

$$\prod(\theta, p \mid \underline{x}) = K p^{n-1} \theta^{-(n+1)} \lambda^{p-1} \exp\left(-\sum_{i=1}^{n} x_i^p / \theta\right)$$

where,

$$K^{-1} = \int_0^\infty \int_0^\infty \prod(\theta, p \mid \underline{x})\, d\theta\, dp$$

$$= \int_0^\infty \int_0^\infty p^{n-1} \lambda^{p-1} \left[\int_0^\infty \theta^{-(n+1)} \exp\left(-\sum_{i=1}^{n} x_i^p / \theta\right) d\theta\right] dp$$

$$= \Gamma(n) \int_0^\infty \frac{p^{n-1} \lambda^{p-1}}{\left\{\sum_{u-1}^{n} x_i^p\right\}^n}\, dp.$$

Thus,

$$\prod(\theta, p \mid \underline{x}) = \frac{p^{n-1} \theta^{-(n+1)} \lambda^{p-1} \exp\left[-\sum_{i=1}^{n} x_i^p / \theta\right]}{\Gamma(n) \int_0^\infty \frac{p^{n-1} \lambda^{p-1}\, dp}{\left(\sum_{i=1}^{n} x_i^p\right)^n}} \tag{2.40}$$

Integrating out θ and p in turn, the marginal posteriors of p and θ are given by:

$$\prod_1 (p \mid \underline{x}) = \frac{\dfrac{p^{n-1}\lambda^{p-1}}{\left(\sum_{i=1}^{n} x_i^p\right)^n}}{\displaystyle\int_0^\infty \frac{p^{n-1}\lambda^{p-1}}{\left(\sum_{i=1}^{n} x_i^p\right)^n}\,dp} \tag{2.41}$$

and, $$\prod_2 (\theta \mid \underline{x}) = \frac{\theta^{-(n+1)}\displaystyle\int_0^\infty \lambda^{p-1} p^{n-1} \exp\left(-\sum_{i=1}^{n} x_i^p/\theta\right) dp}{\Gamma(n)\displaystyle\int_0^\infty \frac{p^{n-1}\lambda^{p-1}\,dp}{\left(\sum x_i^p\right)^n}} \tag{2.42}$$

From (2.41) and (2.42) Bayes estimators are given by:

$$p^* = \frac{\displaystyle\int_0^\infty \frac{p^n\lambda^{p-1}\,dp}{\left(\sum_{i=1}^{n} x_i^p\right)^n}}{\displaystyle\int_0^\infty \frac{p^{n-1}\lambda^{p-1}\,dp}{\left(\sum_{i=1}^{n} x_i^p\right)^n}}, \tag{2.43}$$

and $$\theta^* = \frac{\displaystyle\int_0^\infty \frac{p^{n-1}\lambda^{p-1}}{\left(\sum_{i=1}^{n} x_i^p\right)^{n-1}}\,dp}{(n-1)\displaystyle\int_0^\infty \frac{p^{n-1}\lambda^{p-1}}{\left(\sum_{i=1}^{n} x_i^p\right)^n}\,dp} \tag{2.44}$$

Using the data from Sinha (1986) with $n = 25, \theta = 4, p = 2$, we obtain $p^* = 1.96, \theta^* = 3.65$. (Sinha and Guttman, 1988).

2.9 RAYLEIGH DISTRIBUTION

The Rayleigh pdf of a random variable X is given by

$$f(x \mid \theta) = \frac{x}{\theta^2} \exp\left(-\frac{x^2}{2\,\theta^2}\right), \quad x, \theta > 0. \tag{2.45}$$

Rayleigh distribution is extremely important in communication engineering (Dyer and Whisenand, 1973) and has been successfully used for radiowave power distribution (Siddiqui, 1962) and some types of electrovacuum devices (Polovko, 1968). Rayleigh's model is especially suitable for the life-testing of components or products that age with time.

The likelihood function is

$$L(\theta \mid \underline{x}) \propto \frac{1}{\theta^{2n}} \exp\left(-\frac{S^2}{2\theta^2}\right)$$

where $\underline{x} = (x_1, \cdots, x_n)$ is a random sample from the pdf (2.45) and

$$S^2 = \sum_{i=1}^{n} x_i^2.$$

Consider the natural conjugate family of priors

$$g(\theta) \propto \frac{\exp\left(-\frac{b}{2\theta^2}\right)}{\theta^{a+1}}, \quad a, b > 0 \tag{2.46}$$

The posterior distribution,

$$\Pi(\theta \mid \underline{x}) = \frac{K}{\theta^{2n+a+1}} \exp\left(-\frac{S^2+b}{2\theta^2}\right) \tag{2.47}$$

where, $$K^{-1} = \int_0^\infty \frac{\exp\left(-\frac{S^2+b}{2\theta^2}\right)}{\theta^{2n+a+1}}\, d\theta$$

$$= \frac{1}{2} \int_0^{\infty} \frac{\exp\left(-\frac{S^2+b}{2\theta^2}\right)}{(\theta^2)^{(2n+a)/2+1}} \, d\theta^2$$

$$= \frac{1}{2} \frac{\Gamma\left(\frac{2n+a}{2}\right)}{\left(\frac{S^2+b}{2}\right)^{(2n+a)/2}}.$$

Substituting K in (2.47), we get

$$\Pi(\theta \mid \underline{x}) = \frac{2\left(\frac{S^2+b}{2}\right)^{(2n+a)/2}}{\Gamma\left(\frac{2n+a}{2}\right)} \left\{\frac{\exp\left(-\frac{S^2+b}{2\theta^2}\right)}{\theta^{2n+a+1}}\right\}, \quad \theta, a, b > 0.$$

Bayes estimator,

$$\theta^* = E(\theta \mid \underline{x})$$

$$= K \int_0^{\infty} \frac{\exp\left(-\frac{S^2+b}{2\theta^2}\right)}{\theta^{2n+a}} \, d\theta$$

$$= \frac{K}{2} \int_0^{\infty} \frac{\exp\left(-\frac{S^2+b}{2\theta^2}\right) d\theta^2}{(\theta^2)^{(2n+a-1)/2+1}}$$

$$= \sqrt{\frac{S^2+b}{2}} \, \frac{\Gamma\left(\frac{2n+a-1}{2}\right)}{\Gamma\left(\frac{2n+a}{2}\right)}.$$

2.10 ACCEPTANCE SAMPLING IN THE PRESENCE OF MISCLASSIFICATION

Sinha and Guttman (1994) considered a Bayesian approach to the assessment of the proportion defective in a large lot and estimated the proportion of good items rejected and the proportion of bad items accepted

Suppose a random sample of size n is drawn from a large lot containing an unknown proportion of defective p_0 and let y be the number of defectives in the sample. Further, let $p = P$ (an items is classified defective/the item is defective)

$p' = P$ (an items is classified defective/the item is non-defective)

$z =$ the number of items classified as defective after inspection.

We consider the case where (y, p, p', p_0) are all unknown. All we have is the data z.

Neden (1986) considered the following classification and the corresponding probabilities:

True Classification

True / Classified	non-def.	def.	Total
non-def.	$n-y-z+j$	$y-j$	$n-z$
def.	$z-j$	j	z
Total	$n-y$	y	n

Probabilities Item Classification

True / Classified	non-def.	def.	Total
non-def.	$(1-p_0)(1-p')$	$p_0(1-p)$	$1-p'+p_0p'-p_0p$
def.	$p'(1-p_0)$	pp_0	$p'-p'p_0+pp_0$
Total	$1-p_0$	p_0	1

Let
$$\gamma_{11} = (1-p_0)(1-p'),$$
$$\gamma_{12} = p_0(1-p),$$
$$\gamma_{21} = p'(1-p_0),$$
and
$$\gamma_{22} = 1-\gamma_{11}-\gamma_{12}-\gamma_{21} = pp_0$$

The entries in the Classification table follow a multinomial distribution with probabilities given above. Hence, we have

$$P(y, z, j) = \frac{n!\ \gamma_{11}^{n-y-z+j}\ \gamma_{12}^{y-j}\ \gamma_{21}^{z-j}\ (1-\gamma_{11}-\gamma_{12}-\gamma_{22})^{j}}{j!\,(z-j)!\ (y-j)!\ (n-y-z+j)!}$$

$$= \binom{n}{y}\binom{y}{j}\binom{n-y}{z-j} \gamma_{11}^{n-y-z+j}\, \gamma_{12}^{y-j}\, \gamma_{21}^{z-j}\, (1-\gamma_{11}-\gamma_{12}-\gamma_{21})^{y}$$

where, $0 \le y \le n,\ \ 0 \le z \le n$, (from the margin), and

$0 \le j \le z,\ 0 \le j \le y,\ 0 \le z-j \le z,$

$0 \le z-j \le n-y,\ 0 \le n-y-z+j \le n-z,$

$0 \le n-y-z+j \le n-y,$

$0 \le y-j \le n-z,$

$0 \le y-j \le y$ (from the cell frequencies). (2.48)

To find lower bounds for j, we work from the above conditions and find $0 \le j,\ z+y-n \le j$, and similarly for the upper bounds of j, $j \le z$, $j \le y$ which can be summarized as follows:

$$\max\ (0, z+y-n) \le j \le \min\ (z, y)$$

since the conditions (2.48) holds.

Frequently it is the case that alothough data is collected as summarized by the foregoing tables, only the datum z is recorded, where, once again z is the number of items classified as being defective (out of n), but that inference on (p_0, p', p) is desired. It is a remarkable fact, as this section now describes, that so long as we have an integrable prior for

$$q = P\ (z \le y)$$

we are able *a′ la* Bayes, to make inferences about (p_0, p', p) together, singly or in pairs.

So we consider the following 4 cases:

(i) $\min(y, z) = z$, i.e. $z \le y$ with $z \ge n/2$

(ii) $\min(y, z) = y$, i.e. $y \le z$ with $z \ge n/2$

(iii) $\min(y, z) = z$, i.e. $z \le y$ with $z \le n/2$

(iv) $\min(y, z) = y$, i.e. $y \le z$ with $z \le n/2$

We wlll first analyze the conditions (i) and (ii) which hold for the case $z \ge n/2$ and summarize the results for (iii) and (iv) below, for the case when z is observed to be such that $z \le n/2$.

I. (i) $\min(y, z) = z$, i.e. $z \le y$ with $z \ge \frac{n}{2}$.

$$\text{Let } P_1(z) = \sum_{y=z_j}^{n} \sum_{y+z-n}^{n} P(y, z, j) = \sum\sum_{R_1} \tag{2.49}$$

so that $\sum\sum_{R_1}$ defines the operation $P_1(z)$.

(ii) $\min(y, z) = y$, $y \le z$ with $z \ge \frac{n}{2}$.

Here we are looking at $\max(0, z+y-n) \le j < z$.

Now let

$$P_2(z) = \sum_{j=0}^{n-z} \sum_{y=j}^{n-z} P(y, z, j)$$

$$+ \sum_{y=n-z}^{z} \sum_{j=z+y-n}^{y} P(y, z, j) = \sum\sum_{R_2} \tag{2.50}$$

so that (2.50) defines the operation $P_2(z)$.

Let $q = P(z \le y)$, $1-q = P(y < z)$.

The likelihood function,

$$L(\gamma_{11}, \gamma_{12}, \gamma_{21}, q|z) = q\,P_1(z) + (1-q)\,P_2(z).$$

Bayesian Estimators

Now suppose the prior

$$P(\gamma_{11}, \gamma_{12}, \gamma_{21}, q) = h(q)\;g(\gamma_{11}, \gamma_{12}, \gamma_{21}),$$

where $h(q) = \dfrac{1}{B(a,b)}\, q^{a-1}\,(1-q)^{b-1}$, $0 < q < 1$, $a, b > 0$ (2.51)

and $B(a,b) = \dfrac{\Gamma(a)\,\Gamma(b)}{\Gamma(a+b)}$.

In a state of in-ignorance, Bayesians often assume independence among the parameters of interest (Box and Tiao, 1973). Clearly, in the present set up the assumption of (p_0, p, p') being independent of one another is unacceptable (Biegel, 1974). From the nature of the sampling

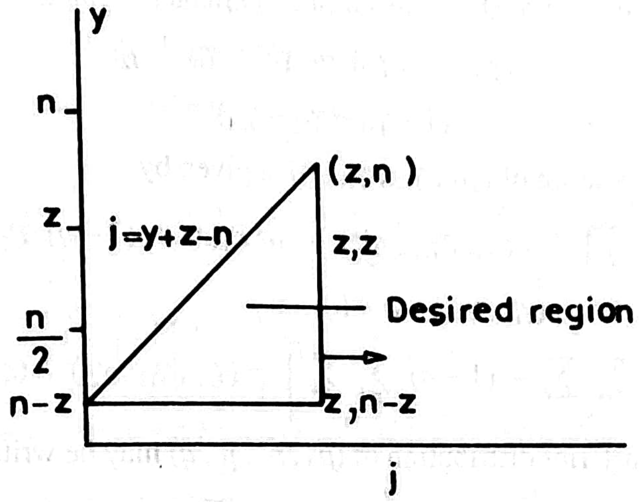

y
n
z
j=y+z-n
(z,n)
z,z
Desired region
n
2
n-z
z,n-z
j

Fig. 2.3

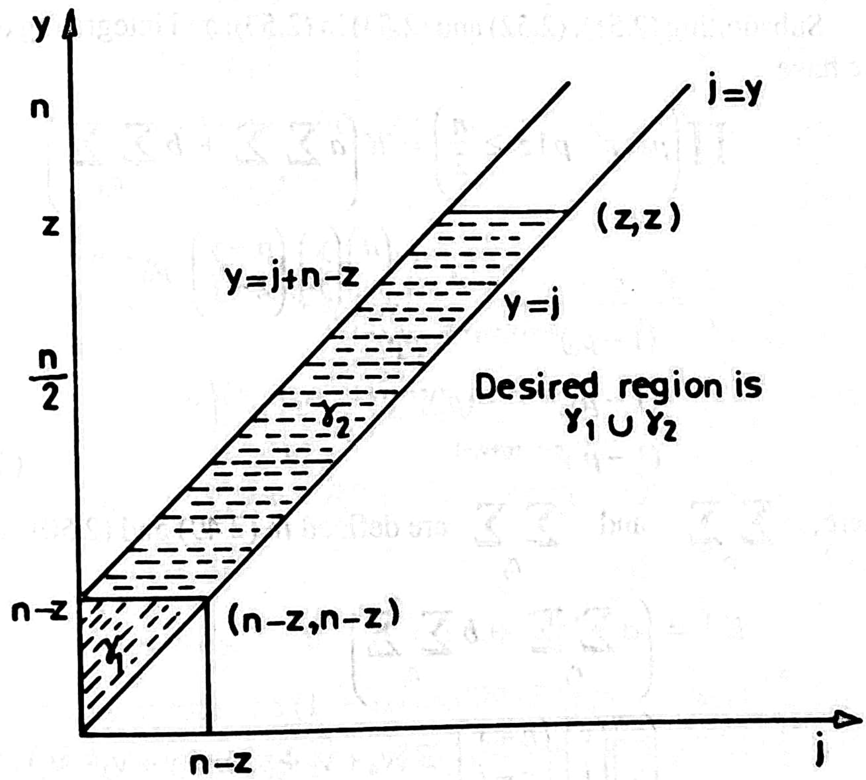

y
n
z
j=y
(z,z)
y=j+n-z
y=j
n
2
Desired region is
γ1 ∪ γ2
γ2
n-z
(n-z,n-z)
γ1
n-z
j

Fig. 2.4

procedure, it appears that the proportion parameters may be highly correlated (Sinclair, 1978). We, therefore, consider Dirichlet's prior distribution,

$$g(\gamma_{11}, \gamma_{21}, \gamma_{12}) \propto \gamma_{11}^{v_1-1}\, \gamma_{12}^{v_2-1}\, \gamma_{21}^{v_3-1}$$
$$(1-\gamma_{11}-\gamma_{12}-\gamma_{21})^{v_4-1} \tag{2.52}$$

The joint posterior of $(\gamma_{11}, \gamma_{12}, \gamma_{21}, q)$ is given by

$$\prod (\gamma_{11}, \gamma_{12}, \gamma_{21}, q|z) \propto [q\, P_1(z) + (1-q)\, P_2(z)]$$
$$g(\gamma_{11}, \gamma_{12}, \gamma_{21})\, h(q)$$
$$\propto \left[q \sum_{R_1}\sum + (1-q) \sum_{R_2}\sum \right] g(\gamma_{11}, \gamma_{12}, \gamma_{21})\, h(q). \tag{2.53}$$

The joint posterior distribution of (p_0, p', p, q) may be written as

$$\prod (p_0, p', p, q \mid z) \propto \prod \gamma_{11}, \gamma_{12}, \gamma_{21}, q|z)\ |J|$$

where
$$|J| = \frac{\partial(\gamma_{11}, \gamma_{12}, \gamma_{21})}{\partial(p_0, p, p')} = p_0(1-p_0). \tag{2.54}$$

Substituting (2.51), (2.52) and (2.54) in (2.53), and integrating out q, we have

$$\prod\left(p_0, p', p \mid z \geq \frac{n}{2}\right) = K\left(a \sum_{R_1}\sum + b \sum_{R_2}\sum\right)$$
$$\binom{n}{y}\binom{y}{j}\binom{n-y}{z-j}\, p_0^{v_4+v_2+y-1}$$
$$(1-p_0)^{n-y+v_3+v_1-1}\ p_1^{v_4+j-1}$$
$$(1-p)^{y-j+v_2-1}\ (p')^{z-v+v_3-1}$$
$$(1-p')^{n-y-z+j+v_1-1} \tag{2.55}$$

where, $\sum_{R_1}\sum$ and $\sum_{R_2}\sum$ are defined in (2.49) and (2.50), and

$$K^{-1} = \left(a \sum_{R_1}\sum + b \sum_{R_2}\sum\right)$$
$$\binom{n}{y}\binom{y}{j}\binom{n-y}{z-j}\, B(v_4+v_2+y, n-y+v_3+v_1).$$
$$B(v_4+j, y-j+v_2)\ B(z-j+v_3,\ n-y-z+j+v_1).$$

Note that $\sum_{R_1}\sum$ and $\sum_{R_2}\sum$ are functions of y and j.

Integrating out p', p from (2.55), the marginal posterior of p_0 given by

$$\Pi_1\left(p_0 \mid z \geq \frac{n}{2}\right) = K\left(a\sum_{R_1}\sum + b\sum_{R_2}\sum\right)$$

$$\binom{n}{y}\binom{y}{j}\binom{n-y}{z-j} B(v_4+j, y-j+v_2)\,.$$

$$B(z-j+v_3, n-y-z+j+v_1)\; p_0^{v_4+v_2+y-1}\,(1-p_0)^{n-y+v_3+v_1-1}\,. \quad (2.56)$$

Similarly, the marginal posteriors of p, p' are given by

$$\Pi_2\left(p \mid z \geq \frac{n}{2}\right) = K\left(a\sum_{R_1}\sum + b\sum_{R_2}\sum\right)$$

$$\binom{n}{y}\binom{y}{j}\binom{n-y}{z-j} B(v_4+v_2+y,$$

$$n-y+v_3+v_1)\; B(z-j+v_3, n-y-z+j+v_1)$$

$$p^{v_4+j-1}\,(1-p)^{y-j+v_2-1} \quad (2.57)$$

and,
$$\Pi_3\left(p' \mid z \geq \frac{n}{2}\right) = K\left(a\sum_{R_1}\sum + b\sum_{R_2}\sum\right)$$

$$\binom{n}{y}\binom{y}{j}\binom{n-y}{z-j} B(v_4+v_2+y,$$

$$n-y+v_1+v_3)\; B(v_3+j, y-j+v_2)\; (p')^{z-j+v_3-1}$$

$$(1-p')^{n-y-z+j+v_1-1} \quad (2.58)$$

$$0 < p_0 < 1\,;\; 0 < p < 1\,;\; 0 < p' < 1\,.$$

We may construct a $(1-\alpha)$ credible interval $(c_1\ c_2)$ for p_0, p, p' from (2.56), (2.57) and (2.58), where c_1 and c_2 satisfy the equations

$$\frac{\alpha}{2} = \int_0^{c_1(p_0)} \Pi_1\left(p_0 \mid z \geq \frac{n}{2}\right) dp_0 = \int_{c_2(p_0)}^{1} \Pi_1\left(p_0 \mid z \geq \frac{n}{2}\right) dp_0 \quad (2.59)$$

$$\frac{\alpha}{2} = \int_0^{c_1(p)} \Pi_2\left(p \mid z \geq \frac{n}{2}\right) dp = \int_{c_2(p)}^{1} \Pi_2\left(p \mid z \geq \frac{n}{2}\right) dp \tag{2.60}$$

and,

$$\frac{\alpha}{2} = \int_0^{c_3(p')} \Pi_3\left(p' \mid z \geq \frac{n}{2}\right) dp' = \int_{c_3(p')}^{1} \Pi_3\left(p' \mid z \geq \frac{n}{2}\right) dp' \tag{2.61}$$

Using squared-error loss function, Bayes estimators of (p_0, p, p') are given by

$$p_0^* = E(p_0 \mid z \geq n/2)$$

$$= K\left(a\sum_{R_1}\sum + b\sum_{R_2}\sum\right)\binom{n}{y}\binom{y}{j}\binom{n-y}{z-j} B(\nu_4+j, y-j+\nu_2)$$

$$B(z-j+\nu_3, n-y-z+j+\nu_1)\ B(\nu_4+\nu_2+y+1, n-y+\nu_1+\nu_3)$$

$$E\left(p_0^2 \mid z \geq \frac{n}{2}\right) = K\left(a\sum_{R_1}\sum + b\sum_{R_2}\sum\right)\binom{n}{y}\binom{y}{j}\binom{n-y}{z-j}$$

$$B(\nu_4+j, -y-j+\nu_2)\ B(z-j+\nu_3, n-y-z+j+\nu_1)$$

$$B(\nu_4+\nu_2+y+2, n-y+\nu_3+\nu_1). \tag{2.62}$$

Hence, $\operatorname{Var}\left(p_0 \mid z \geq \frac{n}{2}\right) = E\left(p_0^2 \mid z \geq \frac{n}{2}\right) - (p_0^*)^2$ may be computed. (2.63)

Similarly,

$$p^* = E\left(p \mid z \geq \frac{n}{2}\right)$$

$$= K\left(a\sum_{R_1}\sum + b\sum_{R_2}\sum\right)\binom{n}{y}\binom{y}{j}\binom{n-y}{z-j}$$

$$B(\nu_4+\nu_2+y, n-y+\nu_1+\nu_3).$$

$$B(z-j+\nu_3, n-y-z+j+\nu_1)$$

$$B(\nu_4+j+1, y-j+\nu_2). \tag{2.64}$$

$$E\left(p^2 \mid z \geq \frac{n}{2}\right) = K\left(a\sum_{R_1}\sum + b\sum_{R_2}\sum\right)$$

$$\binom{n}{y}\binom{y}{j}\binom{n-y}{z-j} B(\nu_4+\nu_2+y\,,n-y+\nu_1+\nu_3)$$
$$B(z-j+\nu_3\,,n-y-z+j+\nu_1)\; B(\nu_4+j+2\,,y-j+\nu_2)\,.$$

and, $$(p')^* = E\left(p' \mid z \geq \frac{n}{2}\right) = K\left(a\sum_{R_1}\sum + b\sum_{R_2}\sum\right)\binom{n}{y}\binom{y}{j}\binom{n-y}{z-j}$$
$$B(\nu_4+\nu_2+y\,,n-y+\nu_1+\nu_3)$$
$$B(\nu_4+j\,,y-j+\nu_2)\; B(z-j+\nu_3+1\,,n-y-z+\nu_1) \quad (2.65)$$

$$E\left(p'^2 \mid z \geq \frac{n}{2}\right) = K\left(a\sum_{R_1}\sum + b\sum_{R_2}\sum\right)\binom{n}{y}\binom{y}{j}\binom{n-y}{z-j}$$
$$B\left(\nu_4+\nu_2+y, n-y+\nu_1+\nu_3\right) B\left(\nu_4+j, y-j+\nu_2\right)$$
$$B\left(z-j+\nu_3+2, n-y-z+\nu_1\right)$$

Thus, $\mathrm{Var}\left(p \mid z \geq \frac{n}{2}\right)$ and $\mathrm{Var}\left(p' \mid z \geq \frac{n}{2}\right)$ may be computed. (2.66)

Putting $\nu_1 = \nu_2 = \nu_3 = \nu_4 = 1$, we obtain Bayes estimator $\{p_0^*, p^*, (p')^*\}$ under the independent ignorance prior $g(\gamma_{11}, \gamma_{12}, \gamma_{21}) =$ constant.

or, $$g(p_0, p, p') \propto p_0(1-p_0)\,.$$

We now consider the case

II. (iii) $\min(y,z) = z\,,\ z \leq y$ with $z \leq \frac{n}{2}$.

Here, $$P_1(z) = \sum_{j=0}^{z}\sum_{y=z}^{j+n-z} P(y,z,j) = \sum_{R_3}\sum \quad (2.67)$$

so that $\sum_{R_3}\sum$ defines the operation $P_1(z)$.

(iv) $\min(y,z) = y$, i.e. $y \leq z$ with $z \leq n/2$. (2.68)

Here, $P_2(z) = \sum_{y=0}^{z}\sum_{j=0}^{y} P(y,z,j) = \sum_{R_4}\sum$, so that $\sum_{R_4}\sum$ defines the operations $P_2(z)$.

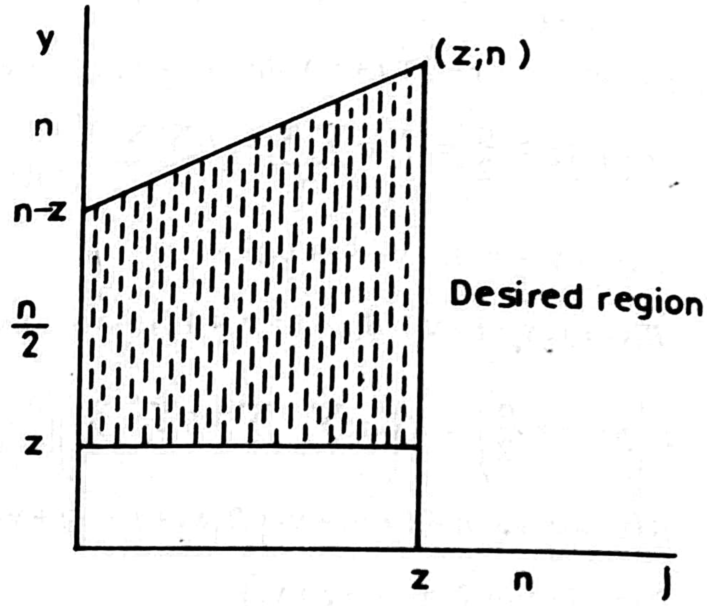


Fig. 2.5

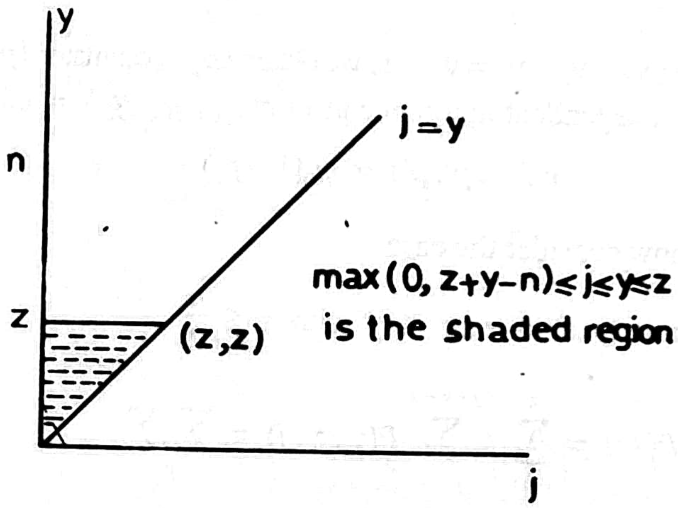


Fig 2.6

All the algebra and computations corresponding to equations (2.56)-(2.66) will be identical except that $\sum\sum_{R_1}$ and $\sum\sum_{R_2}$ will be replaced by $\sum\sum_{R_3}$ and $\sum\sum_{R_4}$, respectively., as defined in (2.67) and (2.68).

Numerical Example

For a given set of (n,z), Bayes estimates $E(p_0 \mid Z)$ $E(p \mid z)$, $E(p' \mid Z)$, and corresponding posterior variances and the $(1-\alpha)$ credible intervals of (p_0, p, p') are computed for $\alpha = 0.01, 0.05$. The results are reported in Tables 2.1–2.6. Entries in the square brackets [ ] represent the left and right tail areas.

Table 2.1: p_0 $(n = 100, z = 20)$

a	b	$E(p_0 \mid Z)$	$V(p_0 \mid Z)$	$(1-\alpha)$ Credible interval $\alpha = 0.01$	$\alpha = 0.05$
1	1	0.49648	0.56141×10^{-1}	(0.04049, 0.95952) [0.00500, 0.00504]	(0.09060, 0.90936) [0.02504, 0.02506]
1	2	0.45520	0.62625×10^{-1}	(0.02482, 0.95037) [0.00331, 0.00663]	(0.05568, 0.88862) [0.01668, 0.03337]
2	1	0.52158	0.50534×10^{-1}	(0.06392, 0.96804) [.00668, .00734]	(0.13825, 0.92538) [0.03171, 0.01821]

Table 2.2: p_0 $(n = 100, z = 80)$

1	1	0.50000	0.56499×10^{-1}	(0.04064, 0.95935) [0.00498, 0.00495]	(0.09104, 0.90862) [0.02498, 0.02507]
1	2	0.47498	0.50809×10^{-1}	(0.03207, 0.93587) [0.00332, 0.00664]	(0.07528, 0.86 [0.01829, 0.03167]
2	1	0.54126	0.63145×10^{-1}	(0.04998, 0.97501) [0.00665, 0.00332]	(0.11174, 0.94391) [0.03324, 0.01674]
1	9	0.45292	0.44756×10^{-1}	(0.02356, 0.86041) [0.00190, 0.00809)	(0.17310, 0.97968) [0.04463, 0.00544]
9	1	0.68144	0.60292×10^{-1}	(0.07827, 0.99130) [0.00896, 0.00100]	(0.17310, 0.97968) [0.04463, 0.00544]

Table 2.3: p $(n = 100, z = 20)$

a	b	$E(p \mid z)$	$V(p \mid Z)$	$(1-\alpha)$ Credible interval $\alpha = 0.01$	$\alpha = 0.05$
1	1	0.26736	0.45700×10^{-1}	(0.0, 0.93391) [0.0, 0.0103]	(0.0, 0.74177) [0.0, 0.04985]
1	2	0.29488	0.55796×10^{-1}	(0.0, 0.96139) [0.0, 0.00974]	(0.0, 0.81754) [0.0, 0.04994]
2	1	0.25063	0.38824×10^{-1}	(0.0, 0.90177) [0.0, 0.00989]	(0.0, 0.67582) [0.0, 0.04991]

Table 2.4: p ($n = 100$, $z = 80$)

a	b				
1	1	0.73425	0.45442×10^{-1}	(0.06300, 1.0) [0.00967, 0.0]	(0.26142, 1.0) [0.05006, 0.0]
1	2	0.72678	0.437512×10^{-1}	(0.05843, 1.0) [0.00956, 0.0]	(0.24816, 1.0) [0.05005, 0.0]
2	1	0.74655	0.41787×10^{-1}	(0.07362, 1.0) [0.00840, 0.0]	(0.28727, 1.0) [0.05022, 0.0]
1	9	0.72020	0.49244×10^{-1}	(0.05813, 1.0) [0.01007, 0.0]	(0.23744, 1.0) [0.05002, 0.0]
9	1	0.78837	0.27103×10^{-1}	(0.12569, 1.0) [0.00998, 0.0]	(0.42439, 1.0) [0.04991, 0.0]

Table 2.5: p' ($n = 100$, $z = 20$)

a	b	$E(p^* \mid z)$	$V(p' \mid Z)$	(1-α) Credible interval	
				$\alpha = 0.01$	$\alpha = 0.05$
1	1	0.26290	0.44303×10^{-1}	(0.0, 0.93030) [0.0, 0.01005]	(0.0, 0.72816) [0.0, 0.05008]
1	2	0.25080	0.40679×10^{-1}	(0.0, 0.92247) [0.0, 0.00994]	(0.0, 0.70348) [0.0, 0.04985]
2	1	0.27025	0.46362×10^{-1}	(0.0, 093637) [0.0, 0.00977]	(0.0, 0.74224) [0.0, 0.04994]

Table 2.6: p' ($n = 100$, $z = 80$)

1	1	0.73425	0.45442×10^{-1}	(0.06300, 1.0) [0.00967, 0.0]	(0.26142, 1.0) [0.05006, 0.0]
1	2	0.75087	0.38583×10^{-1}	(0.09958, 1.0) [0.00991, 0.0]	(0.32667, 1.0) [0.04989, 0.0]
2	1	0.70683	0.55546×10^{-1}	(0.04137, 1.0) [0.01033, 0.0]	(0.18538, 1.0) [0.05030, 0.0]
1	9	0.76552	0.32077×10^{-1}	(0.16580, 1.0) [0.00982, 0.0]	(0.39311, 1.0) [0.05018, 0.0]
9	1	061370	0.78648×10^{-1}	(0.01653, 1.0) [0.00980, 0.0]	(0.08322, 1.0) [0.04915, 0.0]

Discussion

We recall that the prior for q where $q = P(z \leq y)$ given by (2.51) is such that the prior mean $E(q) = \dfrac{a}{a+b}$.

Examining the results of the foregoing Tables we see that for given z, the posterior expectation depends on this prior. For example, (if the prior mean is $\frac{1}{2}$ then we take $a = b = 1$), we find that for $z = 20$, $n = 100$, p_0, the proportion defectives has the posterior with

$$E(p_0 \mid z = 20, n = 100, a = b = 1) = 0.49648 \qquad (2.69)$$

while $$E(p_0 \mid z = 20, n = 100, a = 1, b = 2) = 0.45520 \qquad (2.70)$$

and, $$E(p_0 \mid z = 20, n = 100, a = 2, b = 1) = 0.52158 . \qquad (2.71)$$

The lower result in (2.70) is expected since the prior of q has gone from that located at $\frac{1}{2}$ in (2.69) to $\frac{1}{3}$ in (2.70). We further observe that the result in (2.71) with $E(q) = \frac{2}{3}$ is larger than in (2.69).

In summary, if the prior of q is concentrated below 0.5, the net effect is to "pull" the posterior of p_0 to the left and vice-versa, with similar results for p' (see Tables 2.5 and 2.6). However, it is interesting to note that the "pull" may be in the opposite direction for p (see Tables 2.3 and 2.4).

Table 2.3 exhibits decreasing behaviour as $E(q)$ increases if $z \le \frac{n}{2} (z = 20 < \frac{n}{2} = 50)$, while for the case $z > \frac{n}{2} (z = 80 > 50)$, the behaviour in $E(q)$ is increasing. The authors were surprised by the result for $z \le \frac{n}{2}$ exhibited here which is counter-intuitive.

EXERCISES

1. Let $x = \{x_i\}$. $i = 1, 2, 3, \cdots, n$; $n = 30$ be a random sample from a $N(\mu, \sigma^2)$ and let the joint prior be $g(\mu, \sigma) \propto \frac{1}{\sigma^c}$, $c > 0$.

   (a) Compute (i) Bayes estimate σ^* for $c = 0, 1, 2, 3$ and

   (ii) the raio: $\dfrac{\text{Min}(\sigma^*)}{\text{Max}(\sigma)}$

(b) Plot the posterior distributions $\Pi(\sigma^2 \mid x)$ in (2.6).

Data:

−0.7564	9.8624	10.2535	11.9162	4.0189
4.7247	4.8805	5.4818	7.0530	7.7943
0.4338	1.4677	2.0173	2.5789	4.5249
9.2564	9.9619	10.5209	7.0625	8.0309
4.8404	5.3648	5.7808	3.3948	4.7166
1.3132	1.5292	2.1633	7.2924	8.4281

2. Let μ_1, μ_2, μ_3 be independently normally distributed with prior means $a' = [2,4,8]$,

$$X' = [1,2,4], \quad K = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 1 & \frac{2}{3} \\ \frac{1}{2} & \frac{2}{3} & \frac{3}{4} \end{bmatrix}, \text{ and}$$

$$M = \begin{bmatrix} \frac{1}{4} & \frac{1}{12} & \frac{1}{3} \\ \frac{1}{12} & \frac{1}{4} & \frac{1}{8} \\ \frac{1}{3} & \frac{1}{8} & 1 \end{bmatrix}$$

Find the posterior mean vector $\theta' = [\theta_1, \theta_2, \theta_3]$.

3. Obtain Bayes estimators of μ and σ^2 in the lognormal pdf (2.9) under the joint prior $g(\mu, \sigma) \propto \frac{1}{\sigma^c}$, $c > 0$.

4. Let $\underline{x} = \{x_i\}$; $i = 1, 2, \cdots, n$; $n = 20$ be a random sample from the lognormal pdf (2.9), with mean μ and variance σ^2. Compute Bayes estimates (μ^*, σ^*) under the join prior $g(\mu, \sigma) \propto \frac{1}{\sigma}$.

Data:

16.38	40198.67	26.05	1549.81
0.28	630.81	5.02	43.68
356.43	19.77	1448.26	7.65
16.48	3.78	26.17	15.26
1.09	681.73	6.83	232.40

5. Let $X \sim B(n, p)$ and let V_1 and V_2 be the posterior variances $V(p \mid x)$ under Jeffreys and ALI priors respectively.

   Show that

$$V_1 = \frac{4n^2\,(n+1)\,V_2 + 2n+1}{4(n+1)^2\,(n+2)}\,.$$

6. Obtain Bayes risk for the estimator p^* in (2.18).

7. Show that for the beta-binomial distribution

$$h(x \mid a,b) = \frac{\binom{x+a-1}{a-1}\binom{n-x+b-1}{b-1}}{\binom{n+a+b-1}{a+b-1}}\,,$$

$$x = 0, 1, 2, \cdots, n\,;$$

(i) $\sum_{x=0}^{n} h(x \mid a, b) = 1$, and

(ii) $\text{Var}\,(X \mid a, b) = \dfrac{nab\,(a+b+1)}{(a+b)^2\,(a+b+n)}\,.$

8. Let X ~ Poisson with parameter λ and let V_1, V_2 be the posterior variances $V(\lambda \mid \underline{x})$ with respect to Jeffreys and ALI priors respectively. Show that $V_2 < V_1$.

9. Show that for the estimator λ^* in (2.24), the corresponding Bayes risk is given by

$$R\,(\lambda^*, \lambda) = \frac{a}{b(n+b)}\,.$$

10. Show that for the exponential pdf (2.25), Bayes estimator of θ under the prior $g(\theta) \propto \frac{1}{\theta^3}$ is the same as the classical minimum-mean-square estimator of θ.

11. Show that for the two-parameter exponential pdf (2.27) and the prior $g(\mu, \theta) \propto \frac{1}{\theta^2}$ the Bayes estimators $\{\theta^*, \mu^*\}$ are one and the same as the uniformly-minimum-variance-unbiased estimators

$$\left\{\frac{n\, x_{(1)} - \bar{x}}{n-1}, \quad \frac{n(\bar{x} - x_{(1)})}{n-1}\right\}, \text{ respectively .}$$

12. Derive Bayes estimators $(\theta*, \mu*)$ in the left-truncated exponential distribution (2.32) with respect to the prior $g(\mu, \theta) \propto \frac{1}{\theta^2}$ and compute the estimates, given $S = 15869$, $n = 19$, $x_{(1)} = 162$. (Grubbs, 1971).

13. Suppose the following is a random sample from the two-parameter exponential pdf (2.27). Obtain Bayes estimates (2.30) and (2.31).

Data:

−4.9642	2.4644	5.8441	9.4904	−3.3151
−2.7416	−1.8656	−1.7414	−1.3440	−0.2287
−4.9507	−4.6714	−4.4586	−3.7327	−3.0886
1.4770	3.0191	7.8652	−1.1919	0.0182
−2.4910	−1.7701	−1.6667	−3.3595	−2.8718
−4.9008	−4.4725	−4.3246	−0.8599	0.4455

14. Given the data x as above, plot the marginal posterior's

$$\Pi_1(\mu \mid \underline{x}) \text{ and } \Pi_2(\theta \mid \underline{x}) \text{ in (2.29) and (2.30).}$$

15. Suppose $x = \{x_i\}$, $i = 1, 2, 3, \cdots, n$; $n = 30$ is a random sample from the left-truncated exponential pdf (2.26). Compute the estimates $\mu^*, \theta*)$ in (2.34) and (2.36).

Data:

5.0502	15.7113	18.3002	21.6188	6.7114
6.9183	7.3420	7.8805	9.0407	10.9472
5.1784	5.4052	5.6505	5.6699	6.7663
14.4270	17.5888	18.9296	9.9947	11.9806
7.0247	7.3631	8.0558	6.7006	6.7879
5.2000	5.4416	5.6578	10.4296	12.4674

16. Plot the marginal posteriors $\Pi_1 (\mu \mid \underline{x})$ and $\Pi_2 (\theta \mid \underline{x})$ in (2.33) and (2.36) using the data x in the preceding exercise.

17. Let $\underline{x} = \{x_i\}$, $i = 1, 2, , \cdots , n$; $n = 30$ be a random sample from the Weibull pdf (2.30). Compute Bayes estimates (θ^*, p^*) in (2.43) and (2.44).

Data:

5.0077	11.6369	13.8027	19.4158	5.6762
5.9258	6.3500	6.9005	8.1383	9.3443
5.1095	5.3508	5.4352	5.5131	5.7342
10.8623	12.7530	14.0166	8.6077	9.6105
6.3226	6.4556	7.3815	5.6047	5.8241
5.1855	5.4325	5.4464	9.2234	10.3455

18. Let the following data $\underline{x}$ be a random sample from the Rayleigh pdf (2.45). Compute Bayes estimates of θ with respect to ALI and Jeffreys prior for θ.

Data:

1.8637	9.7849	10.8185	6.3818	7.6122
4.9225	5.5530	6.0044	3.7891	4.5566
2.5716	3.0413	3.4862	6.6899	7.6331
9.7437	9.8272	12.7716	3.9870	4.7411
5.0859	5.8119	6.1669	6.8934	9.6217
2.8264	3.4174	3.6742	4.5439	15.9188

19. Let V_1 and V_2 be $V(\theta^2 \mid \underline{x})$ for the Rayleigh pdf (2.45) under Jeffreys and ALI priors $g_1(\theta)$ and $g^2(\theta)$, repsectively. Show that $V_1 > V_2$.

20. Consider the posterior $\Pi(\theta \mid \underset{\sim}{x})$ in (2.47). Show that $\theta^{2*} = E(\theta^2 \mid \underset{\sim}{x}) = \dfrac{S^2 + b}{(2n + a - 2)}$ and the associated Bayes risk

$$R^*(\theta^{*^2}, \theta) = \frac{2b^2}{(2n + a - 2)\,(a - 2)\,(a - 4)}, \quad a > 4.$$

21. Obtain Bayes risk $R^*(\theta^*, \theta)$ of the estimator $\theta^* = \dfrac{S + a}{n + b - 1}$ as derived in the section 2.7 for the exponential distribution.

22. Let $X \sim N(\theta, 1)$ and $\theta \sim N(\mu, 1)$.

Obtain $E(\theta \mid \underset{\sim}{x})$, $\underset{\sim}{x} = (x_1, x_2, \cdots, x_n)$.

CHAPTER

3 Bayesian Predictive Distribution and Prediction Interval

3.1 INTRODUCTION

Given sufficient information about the past and present behaviour of an event or an observation, an important objective in scientific investigation is to predict the nature of its future behaviour. Predictive distributions for a number of failure time distributions have been studied by Aitchison (1964), Sinha (1985) among others.

Suppose n units have been subjected to some life testing experiment. Let the failure times $\underline{x} = (x_1, x_2, \cdots x_n)$ be a random sample from some distribution $F(x \mid \theta)$ characterized by a density function $f(x \mid \theta)$. Suppose m units of the same kind are to be put into future use and let the future failure times $\underline{y} = (y_1, y_2, \cdots y_m)$ be the second independent sample from the same distribution $F(y \mid \theta)$. In life testing and quality control problems it is often important to make prediction about some function of $(y_1, y_2, \dots y_m)$ on the basis of the previously obtained life test data $\underline{x}$.

The Bayesian predictive distribution of a future observation Y is defined as the posterior expectation of $f(y \mid \theta)$, and is given by

$$h(y \mid \underline{x}) = C \int_{\Omega} f(y \mid \theta)\, \Pi(\theta \mid \underline{x})\, d\theta \tag{3.1}$$

where C is the normalizing constant, $\Pi(\theta \mid \underline{x})$ is the posterior distribution θ and Ω is the range space of θ.

We now consider the prediction limits and prediction interval for a univariate distribution. Prediction interval would be a straightforward and natural consequence of Predictive Distribution we have defined above. Forecasting the next observation may be very important in a certain process. The government may wish to predict the total revenue income for the next fiscal year from a historical record of revenue accumulation to formulate the taxation policy or allocation of resources for various projects. Whitmore (1986), Nelson (1982) and Dudewicz (1976) have illustrated the application of prediction limits in city government industry and quality assurance.

$(1-\alpha)$-equal-tail-prediction interval (L, U) of a future observation Y is defined by $1-\alpha = P[L < Y < U]$

$$= \int_L^U h(y \mid \underline{x})\, dy, \text{ where } \frac{\alpha}{2} = \int_{-\infty}^{L} h(y \mid \underline{x})\, dy = \int_U^{\infty} h(y \mid \underline{x})\, dy \quad (3.2)$$

and L and U are the corresponding lower and upper $(1-\alpha)$ predictive limits of Y.

3.2 EXPONENTIAL DISTRIBUTIONS

Consider the one-parameter exponential pdf

$$f(x \mid \theta) = \frac{1}{\theta} \exp\left(-\frac{x}{\theta}\right), x, \theta > 0 \quad (3.3)$$

and Jeffreys (1961) prior

$$g(\theta) \propto \frac{1}{\theta}.$$

From (2.21) the posterior distribution

$$\Pi(\theta \mid \underline{x}) \propto \frac{\exp\left(-\frac{S}{\theta}\right)}{\theta^{n+1}}, \quad S = \sum_{i=1}^{n} x_i \quad (3.4)$$

From (3.1) and (3.3), the predictive distribution

$$h(y \mid \underline{x}) \propto \int_0^{\infty} \frac{\exp\left(-\frac{S+y}{\theta}\right)}{\theta^{n+2}} d\theta$$

$$= \frac{C}{(S+y)^{n+1}},$$

where $$C^{-1} = \int_0^{\infty} \frac{dy}{(S+y)^{n+1}} = \frac{1}{n S^n}$$

Restoring the normalizing constant

$$h(y \mid \underline{x}) = \frac{n S^n}{(S+y)^{n+1}} \quad y > 0 \tag{3.5}$$

which we write as

$$h_B(w) = \frac{n}{(1+w)^{n+1}} \quad w > 0$$

where $W = \frac{Y}{S}$ and $h_B(w)$ represents the pdf of W using Bayesian frame work.

Consider the classical predictive distribution of Y.

$$h(y \mid \theta) = \frac{1}{\theta} \exp\left(-\frac{y}{\theta}\right)$$

Let $$\frac{Y}{\theta} = U, \frac{S}{\theta} = V, W = \frac{Y}{S} = \frac{U}{V}.$$

$$F(w) = P(W \le w)$$

$$= P(U \le v w).$$

It is well-known that $V \sim \gamma(n)$ (Appendix A.3)

Hence, $$F(w) = \frac{1}{\Gamma(n)} \int_0^{\infty} v^{n-1} \exp(-v) \left[\int_0^{vw} \exp(-u)\, du\right] dv$$

$$= \frac{1}{\Gamma(n)} \int_0^{\infty} v^{n-1} \exp(-v)\, [1 - \exp(-v w)]\, dv$$

$$= 1 - \frac{1}{\Gamma(n)} \int_0^{\infty} v^{n-1} \exp \{-v(1+w)\} dv$$

$$= 1 - \frac{1}{(1+w)^n}$$

$$h_C(w) = \frac{n}{(1+w)^{n+1}}, \quad w > 0 \tag{3.6}$$

Thus, the classical and Bayesian predictive distribution of a future observation (with respect to the prior $g(\theta) \propto \frac{1}{\theta}$) are one and the same and hence, the corresponding prediction intervals will be the same as well.

Note that for the numerical computation of the prediction interval, the Bayesian route is usually simpler.

Let a and b be the lower and upper Bayesian $(1-\alpha)$ predictive limits of a future observation Y. From (3.5) we have

$$\frac{\alpha}{2} = \int_0^a \frac{n S^n}{(S+y)^{n+1}} dy = \int_b^{\infty} \frac{n S^n}{(S+y)^{n+1}} dy$$

which leads to $$a = \left[\left(1 - \frac{\alpha}{2}\right)^{-\frac{1}{n}} - 1\right] S,$$

and $$b = \left[\left(\frac{\alpha}{2}\right)^{-\frac{1}{n}} - 1\right] S.$$

For the classical part, from (3.6) it follows that

$$\frac{Y}{S} \sim \frac{1}{n} F(2, 2n)$$

$$1 - \alpha = P[F_1 < F < F_2]$$

$$= P\left[F_1 < \frac{ny}{S} < F_2\right]$$

$$= P\left[\frac{SF_1}{n} < y < \frac{SF_2}{n}\right].$$

where F_1 and F_2 are lower and upper $\left(\frac{\alpha}{2}\right)$ - points of an F-distribution with 2 and $2n$ degrees of freedom. Thus, the classical limits $(y^L, y^U) \equiv \left(\frac{SF_1}{n}\ \frac{SF_2}{n}\right)$ require the use of the F-table while the Bayesian limits (a, b) are given (3.8) in simple and closed forms.

Let $\{x_{(1)}, x_{(2)}, \ldots, x_{(n)}\}$ be the recorded failure times of n components of a system and let $\{y_{(1)}, y_{(2)}, \ldots, y_{(m)}\}$ be the failure times of a future sample of m similar components. Given $\{x_{(1)}, x_{(2)}, \ldots x_{(n)}\}$ we are interested in the predictive distribution and the prediction interval of the k^{th} failure time $Y_{(k)}$. For simplicity we will write $y_{(i)} = y_i$ and similarly for $x_{(i)}$ and $\underline{x} = (x_1, x_2, \ldots x_n)$. Suppose the underlying failure-time distribution is the exponential pdf (3.3). Let Y_k be the k^{th} failure time in a future sample $\underline{y} = (y_1, y_2, \cdots, y_m)$ from the exponential pdf,

$$f(y \mid \theta) = \frac{1}{\theta} \exp\left(-\frac{y}{\theta}\right), \quad y, \theta > 0.$$

We have,

$$h(y_{(k)} \mid \underline{x}) = \int_0^\infty f(y_{(k)} \mid \theta)\, \Pi(\theta \mid \underline{x})\, d\theta. \tag{3.8}$$

For simplicity we will write $Y_{(k)} = Y_k$.

$$f(y_k \mid \theta) \propto \theta^{-1}\left\{1 - \exp\left(-\frac{y_k}{\theta}\right)\right\}^{k-1} \exp\left\{-\frac{y_k(m-k+1)}{\theta}\right\}$$

$$\propto \sum_{i=0}^{k-1} \frac{(-1)^i}{\theta} \binom{k-1}{i} \exp\left\{-\frac{y_k(m-k+1+i)}{\theta}\right\}.$$

From (3.4) and (3.8), the Bayesian predictive distribution of $Y_{(k)}$ is given by

$$h(y \mid \underline{x}) \propto \sum_{i=0}^{k-1} (-1)^i \binom{k-1}{i} \int_0^\infty \frac{\exp\left[-\frac{1}{\theta}\{y_k(m-k+1+i) + S\}\right]}{\theta^{n+2}}\, d\theta$$

$$\propto \sum_{i=0}^{k-1} (-1)^i \binom{k-1}{i} [y_k(m-k+1+i) + S]^{-(n+1)}$$

$$\doteq A \sum_{i=0}^{k-1} (-1)^i \binom{k-1}{i} [y_k (m-k+1+i)+S]^{-(n+1)},$$

$$\text{where, } A^{-1} = \sum_{i=0}^{k-l} (-1)^i \binom{k-1}{i} \int_0^{\infty} \frac{d\, y_k}{\{y_k (m-k+1+i)+S\}^{n+1}}$$

$$= \sum_{i=0}^{k-1} \frac{(-1)^i \binom{k-1}{i}}{m-k+1+i} \int_s^{\infty} \frac{d\nu}{\nu^{n+1}}$$

$$= \frac{1}{nS^n} \sum_{i=0}^{k-1} \frac{(-1)^i \binom{k-1}{i}}{m-k+1+i} = \frac{B(k, m-k+1)}{nS^n} \quad \text{(Appendix A.4)}$$

Thus,

$$h(y_k \mid \underline{x}) = \frac{n S^n}{B(k, m-k+1)} \sum_{i=0}^{k-1} \binom{k-1}{i}$$

$$(-1)^i [y_k(m-k+1+i)+S]^{-(n+1)} \quad y_k > 0. \tag{3.10}$$

Substituting $\frac{Y_k}{S} = W$ in (3.10), we write

$$g_B(w) = \frac{n}{B(k, m-k+1)} \sum_{i=0}^{k-1} \binom{k-1}{i}$$

$$(-1)^i [(m-k+1+i)\, w+1]^{-(n+1)} \quad w > 0. \tag{3.11}$$

We now derive the classical counterpart of (3.11)

$$h(y_k \mid \theta) = \frac{k \binom{m}{k}}{\theta} \left\{1 - \exp\left(-\frac{y_k}{\theta}\right)\right\}^{(k-1)}$$

$$\exp\left\{-\frac{y_k}{\theta}(m-k+1)\right\}$$

$$= \frac{k\binom{m}{k}}{\theta} \sum_{i=0}^{k-1} (-1)^i \binom{k-1}{i} \cdot \exp\left\{\frac{-\, y_k(m-k+1+i)}{\theta}\right\}. \tag{3.12}$$

Let, $\dfrac{Y_k}{\theta} = U, \dfrac{S}{\theta} = V, W = \dfrac{U}{V} = \dfrac{Y_k}{S}$.

From (3.12) we have,

$$h(u) = k\binom{m}{k}\sum_{i=0}^{k-1}(-1)^i\binom{k-1}{i}\exp\{-u(m-k+1+i)\}$$

$$F(w) = P(W \le w)$$

$$= P(U \le v\,w)$$

$$= \frac{1}{\Gamma(n)}\int_0^{\infty} v^{n-1}\exp(-v)\left[\int_0^{vw} h(u)\,du\right]dv$$

$$= \frac{k\binom{m}{k}}{\Gamma(n)}\sum_{i=0}^{k-1}(-1)^i\binom{k-1}{i}(m-k+1+i)^{-1}$$

$$\int_0^{\infty} v^{n-1}\exp(-v)\,[1-\exp\{-v\,w(m-k+1+i)\}]\,dv$$

$$= k\binom{m}{k}\sum_{i=0}^{k-1}(-1)^i\binom{k-1}{i}$$

$$(m-k+1+i)^{-1}\,[1-\{1+w(m-k+1+i)\}^{-n}]$$

$$g_C(w) = n\,k\binom{m}{k}\sum_{i=0}^{k-1}(-1)^i\binom{k-1}{i}[1+w(m-k+1+i)]^{-(n+1)}$$

$$= \frac{n}{B(k, m-k+1)}\sum_{i=0}^{k-1}(-1)^i\binom{k-1}{i},$$

$$[1+w(m-k+1+i)]^{-(n+1)}, \quad w > 0 \tag{3.13}$$

$$= g_B(w) \text{ in } (3.11).$$

Thus, the Bayesian and classical predictive distributions of $Y_{(k)}$ are one and the same and, hence, their corresponding prediction intervals will be identical.

By putting $m = k = 1$ in (3.10) and (3.13), we obtain (3.5) and (3.6).

From (3.10) it follows that the $(1 - \alpha)$-prediction limits (a, b) of Y_k are solutions of

$$\frac{\alpha}{2} = \int_0^a h\,(y_k \mid \underline{x})\,d\,y_k = \int_b^\infty h\,(y_k \mid \underline{x})\,d\,y_k$$

which leads to the following equations:

$$\frac{\alpha\,B\,(k\,,m-k+1)}{2\,n\,S^n} = \sum_{i=0}^{k-1} (-1)^i \binom{k-1}{i}$$

$$(m-k+1+i)^{-1}\;[S^n - \;\{a\,(m-k+1+i)+S\;\}^{-n}]\,.$$

$$\frac{\alpha\,B\,(k\,,m-k+1)}{2\,n\,S^n} = \sum_{i=0}^{k-1} \binom{k-1}{i}$$

$$(m-k+1+i)^{-1}\;[S+b\;\{a(m-k+1+i)\;]^{-n}\,.$$

Putting $m = k = 1$, a and b are given by

$$a = \left[\left(1 - \frac{\alpha}{2}\right)^{-\frac{1}{n}} - 1\right] S\,,$$

and

$$b = \left[\left(\frac{\alpha}{2}\right)^{-\frac{1}{n}} - 1\right] S \quad \text{obtained earlier}\,.$$

Consider the two-parameter exponential pdf

$$f(x \mid \mu\,,\theta) = \frac{1}{\theta}\exp\left(-\frac{x-\mu}{\theta}\right), -\infty < \mu\,, < x_{(1)} < \infty\,,\,\theta > 0\,. \tag{3.14}$$

Let $Y_{(k)} = Y_k$ be the k^{th} order statistic in a future sample $\underline{Y} = (y_1\,,y_2\,,\;\cdots\;,y_m)$ from the pdf (3.14). From (2.28) we have the joint posterior distribution of μ and θ given by

$$\prod(\mu\,,\theta \mid \underline{x}) \propto \frac{1}{\theta^{n+1}}\exp\left[-\frac{S+n\,(x_{(1)}-\mu\,)}{\theta}\right] \tag{3.15}$$

$$f(y_k \mid \mu\,,\theta) \propto \frac{1}{\theta}\left[1 - \exp\left(-\frac{y_k-\mu}{\theta}\right)\right]^{k-1}$$

$$\exp\left[-\frac{1}{\theta}\left\{(y_k-\mu)(m-k+1)\right\}\right]$$

$$\propto \sum_{i=0}^{k-1}(-1)^i\binom{k-1}{i}\exp\left[-\frac{1}{\theta}\left\{(y_k-\mu)\ (m+k+1+i)\right\}\right]. \tag{3.16}$$

From (3.1), (3.15) and (3.16) we have the predictive distribution of Y_k given by

$$h\,(y_k\,|\,\underline{x}) \propto \int\int_0^{\infty}\sum_{i=0}^{k-1}\binom{k-1}{i}(-1)^i\frac{1}{\theta^{n+2}}\exp\left[-\frac{1}{\theta}\right.$$

$$\left.\left\{S+n\,x_{(1)}+y_k\,(m-k+1+i)-\mu\,(m-k+1+i+n)\right\}\right]d\theta\,d\mu$$

$$\propto \sum_{i=0}^{k-1}(-1)^i\binom{k-1}{i}\int_{-\infty}^{x_{(1)}}[S+n\,x_{(1)}+y_k\,(m-k+1+i)$$

$$-\mu\,(m-k+1+i+n)]^{-(n+1)}\,d\mu. \tag{3.17}$$

We have to consider two cases:

(i) $y_k > x_{(1)}$.

$$h\,(y_k\,|\,\underline{x}) \propto \sum_{i=0}^{k-1}(-1)^i\binom{k-1}{i}\int_{-\infty}^{x_{(1)}}[S+n\,x_{(1)}+y_k\,(m-k+1+i)$$

$$-\mu\,(m-k+n+1+i)]^{-(n+1)}\,d\mu$$

$$= A\sum_{i=0}^{k-1}(-1)^i\binom{k-1}{i}(m-k+i+n+1)^{-1}$$

$$[S+(y_k-x_{(1)})\ (m-k+1+i)]^{-n}$$

where A is the normalizing constant.

(ii) $y_k < x_{(1)}$.

We have,

$$h(y_k \mid \underset{\sim}{x}) \propto \sum_{i=0}^{k-1} (-1)^i \binom{k-1}{i} \int_{-\infty}^{y_k} [S + n\,x_{(1)} + y_k\,(m-k+1+i)$$

$$-\mu\,(m-k+1+n+i)]^{-(n+1)}\, d\mu$$

$$= A\,\{S + n\,(x_{(1)} - y_k)\,\}^{-n} \sum_{i=0}^{k-1} (-1)^i$$

$$\binom{k-1}{i} / (m-k+1+i+n),$$

where

$$A^{-1} = \sum_{i=0}^{k-1} (-1)^i \binom{k-1}{i} (m-k+1+i+n)^{-1}]$$

$$\left[\int_{x_{(1)}}^{\infty} \{S + (m-k+1+i)\,(y_k - x_{(1)})\,\}^{-n}\, d\,y_k + \int_{-\infty}^{x_{(1)}} \{S + n\,(x_{(1)} - y_k)\,\}^{-n}\, d\,y_k\right]$$

$$= B\,(k, m-k+1) / n\,(n-1)\,S^{n-1}$$

$$= \frac{1}{n\,(n-1)\,k \binom{m}{k} S^{n-1}}.$$

Thus,

$$h(y_K \mid \underset{\sim}{x}) = n\,(n-1)\,k \binom{m}{k} S^{n-1} \sum_{i=0}^{k-1}$$

$$(-1)^i \binom{k-1}{i} (m-k+1+i+n)^{-1}$$

$$\{S + (y_k - x_{(1)})\ (m-k+1+i)\ \}^{-n} \quad y_k \geq x_{(i)} \equiv h_2 \quad (3.18)$$

and,

$$= n\,(n-1)\,k \binom{m}{k} S^{n-1} \quad \{S + n\,(x_{(1)} - y_k)\,\}^{-n}$$

$$\sum_{i=0}^{k-1} (-1)^i \binom{k-1}{i} (m-k+1+i+n)^{-1} \qquad \text{(Appendix A.5)}$$

$$= \frac{n\,(n-1)\,S^{n-1}\,m^{(k)}}{(m+n)^{(k)}} \;\{(S+n\,(x_{(1)}-y_k)\,\}^{-n} \quad y_k \le x_{(1)}$$

$$\equiv h_1 \tag{3.19}$$

where $\quad m^{(\gamma)} = m\,(m-1)\,(m-2)\cdots(m-\gamma+1)\,.$

Substituting $\dfrac{n\,(y_k - x_{(1)})}{S} = z$ in (3.18) and (3.19) we obtain

$$h_B(z) = (n-1)\,k\binom{m}{k}\sum_{i=0}^{k-1}(-1)^i\binom{k-1}{i}(m-k+1+i+n)^{-1}$$

$$\left[1+\frac{z}{n}(m-k+1+i)\right]^{n},\; z > 0 \tag{3.20}$$

and $\qquad = (n-1)\;m^{(k)}\;/\;(m+n)^{(k)}\;(1-z)^n,\, z < 0 \qquad (3.21)$

which is the same as $h_C(z)$, the density function of Z as obtained by Lawless (1977) using the classical approach.

Thus, $\qquad h_B(z) \equiv h_C(z)$

and so are the predictive limits of $y_{(k)}$.

Let (L, U) be the classical/Bayesian $(1-a)$-predictive interval of $Y_{(k)}$. L and U are solutions of

$$\int_{-\infty}^{L} h_1(y_k \mid \underset{\sim}{x})\,d\,y_k = \frac{\alpha}{2} = \int_{U}^{\infty} h_2(y_k \mid \underset{\sim}{x})\,d\,y_k \tag{3.22}$$

where $h_1(y_k \mid \underset{\sim}{x})$ and $h_2(y_k \mid \underset{\sim}{x})$ are given by (3.19) and (3.18) respectively, which leads to

$$\frac{\alpha\,(m+n)^{(k)}}{2\,m^{(k)}\,S^{n-1}} = \{S+n\,(x_1-L)\,\}^{-(n-1)},$$

and $$\frac{\alpha}{2\,n\,k\binom{m}{k}S^{n-1}} = \sum_{i=0}^{k-1}(-1)^i\binom{k-1}{i}\{(m-k+1+i+n)$$

$$(m-k+1+i)\,\}^{-1}\;\;\{S+(m-k+1+i)\;(U-x_1)\,\}^{-(n-1)}$$

From $m = k = 1$, L and U are solutions of

$$\frac{\alpha(n+1)}{2\,S^{n-1}} = \{S+n\,(x_1 - L)\ \}^{-(n-1)}, \text{ and}$$

$$\frac{\alpha(n+1)}{2\,n\,S^{n-1}} = \{S+U-x_1\ \}^{-(n-1)}.$$

Now suppose Y_K is the k^{th} order statistic in a future sample $\underline{y} = (y_1, y_2, \cdots, y_m)$ from the left-truncated exponential pdf

$$f(x \mid \mu, \theta) = \frac{1}{\theta}\exp\left(-\frac{x-\mu}{\theta}\right),\ 0 < \mu < x_{(1)} < \infty,\ \theta > 0. \quad (3.23)$$

Here the range of integrating with respect to μ will change in (3.17).

(i) $y_k > x_{(1)}$:

$$h(y_k \mid \underline{x}) \propto \sum_{i=0}^{k-1} (-1)^i \binom{k-1}{i} \int_0^{x_{(1)}} [S + n\,x_{(1)} +$$

$$y_k\,(m-k+1+i) - \mu\,(m-k+1+i+n)]^{-(n+1)}\,d\mu$$

$$= C \sum_{i=0}^{k-1} (-1)^i \binom{k-1}{i} \cdot$$

$$(m-k+1+i+n)^{-1}\,[\{S+(m-k+1+i)\cdot$$

$$(y_k - x_{(1)})\ \}^{-n} - \{S+n\,x_{(1)} + y_k\,(m+k+1+i)\ \}^{-n}]\,.$$

(ii) $y_k < x_{(1)}$:

$$h(y_k \mid \underline{x}) \propto \sum_{i=0}^{k-1} (-1)^i \binom{k-1}{i} \int_0^{y_{(k)}} [S+n\,x_{(1)} + y_k\,(m-k+1+i)$$

$$-\mu\,(m-k+1+i+n)]^{-(n+1)}\,d\mu$$

$$= C \sum_{i=0}^{k-1} (-1)^i \binom{k-1}{i}\ .$$

$$(m-k+1+i+n)^{-1}\ [\ \{S+n\,(x_{(1)} - y_k)\}^{-n}$$

$$-\{S+n\,x_{(1)} + y_k\ (m+k+1+i)\ \}^{-n}],$$

where, $$C^{-1} = \sum_{i=0}^{k-1} (-1)^i \binom{k-1}{i} (m-k+1+i+n)^{-1}$$

$$+ \left[\int_0^{x_{(1)}} \{S + n\,(x_{(1)} - y_k)\}^{-n}\, d\,y_k + \int_{x_{(1)}}^{\infty} \{S + (m-k+1+i)\right.$$

$$\left. - \;(y_k - x_{(1)})\}^{-n} d\,y_k + \int_0^{-\infty} \{S + n\,x_{(1)} + y_k\,(m-k+1+i)\}^{-n}\; d\,y_k \right]$$

$$= \frac{1}{n\,(n-1)\,S^{n-1}} \left\{1 - \left(\frac{S}{S+n\,x_{(1)}}\right)^{n-1}\right\} \sum_{i=1}^{k-1} \frac{(-1)^i \binom{k-1}{i}}{(m-k+1+i)}$$

$$= \frac{1}{n\,(n-1)\,k \binom{m}{k} S^{n-1}} \left\{1 - \left(\frac{S}{S+n\,x_{(1)}}\right)^{n-1}\right\}$$

$$C = n\,(n-1)\,k \binom{m}{k} S^{n-1} \left\{1 - \left(\frac{S}{S+n\,x_{(1)}}\right)^{n-1}\right\}^{-1}. \tag{3.24}$$

Thus,

$$h\,(y_k \mid \underline{x}) = C \sum_{i=0}^{k-1} (-1)^i \binom{k-1}{i} (m-k+1+i+n)^{-1} \cdot$$

$$\left[\{S + (m-k+1+i)\;(y_k - x_{(1)})\}^{-n} \right.$$

$$\left. - \;\{S + n\,x_{(1)} + y_k\,(m-k+1+i)\}^{-n}\right],\; y_k > x_{(1)}$$

$$\equiv h_2 \tag{3.25}$$

and,

$$= C \sum_{i=0}^{k-1} (-1)^i \binom{k-1}{i} (m-k+i+1+n)^{-1}$$

$$[\; \{S + n(x_{(1)} - y_k)\}^{-n}$$

$$-\; \{S + n\,x_{(1)} + y_k\,(m-k+1+i)\}^{-n}],\; y_k < x_{(1)}$$

$$\equiv h_1 \tag{3.26}$$

where C is given by (3.24).

The lower and upper $(1-\alpha)$-prediction limits L and U are solutions of the equations

$$\frac{\alpha}{2} = \int_0^L h_1\, d\, y_k = \int_U^\infty h_2\, d\, y_k$$

where h_1 and h_2 are defined in (3.26) and (3.25), which leads to the closed form equations,

$$\alpha\, n\,(n-1)\,(2B_1)^{-1} + B\,(k\,,m-k+1)\ \{n\ \bar{x}\}^{-(n-1)}$$

$$= B\,(k\,,m-k+1+n)\ \{n(\bar{x}-L)\ \}^{-(n-1)}$$

$$+\ n \sum_{i=0}^{k-1} (-1)^i . \binom{k-1}{i}\ [\ \{(m-k+1+i)\ (m-k+1+i+n)\ \}^{-1}$$

$$\{n\,\bar{x} + L\,(m-k+1+i)\ \}^{-(n-1)}] \qquad (3.27)$$

and,

$$\alpha\,(n-1)\,(2\,B_1)^{-1} = \sum_{i=0}^{k-1} (-1)^i \binom{k-1}{i}$$

$$\{(m-k+1+i)\cdot(m-k+1+i+n)\ \}^{-1}\cdot$$

$$[\{s+(m-k+1+i)(U-x_1)\ \}^{-(n-1)}$$

$$-\ \{n\,\bar{x}+(m-k+1+i)\,U\ \}]^{-(n-1)}. \qquad (3.28)$$

For $m = k = 1$. L and U are solutions of

$$\{\alpha\, n\,(n-1)\ \ (n+1)\}\ \ (2B_1)^{-1} + (n+1)\,(n\,\bar{x})^{-(n-1)}$$

$$=\ \ \{n\,(\bar{x}-L)\ \}^{-(n-1)} + n\,(n\,\bar{x}+L)^{-(n-1)}$$

and,

$$\{\alpha\, n(n-1)\,(n+1)\}\ (2\,B_1)^{-1} = \{s+U-x_1\}^{-(n-1)} - \{n\,\bar{x}+U\ \}^{-(n-1)}.$$

Classical Approach

Evans and Nigm (1980) remarked that it would be extremely difficult for a classical approach to prediction incorporating the prior information $\mu \geq 0$. We obtain such a distribution in the following. The joint density,

$$f(x_1, y_k \mid \mu, \theta, \underset{\sim}{x}) \propto \left[1 - \exp\left\{-\frac{(y_k - \mu)}{\theta}\right\}\right]^{k-1} \cdot$$

$$\exp\left[-\left\{\frac{(m-k+1)\ (y_k - \mu)}{\theta}\right\}\right] \cdot \exp\left[-\ \left\{\frac{n\,(x_1 - \mu)}{\theta}\right\}\right]$$

Let $\begin{matrix} y_k - x_1 = t_1 & y_k = t_1 + t_2 > \mu \\ x_1 = t_2 & x_1 = t_2 > \mu \end{matrix}$, which implies

$t_2 > \mu - t_1$ and also $t_2 > \mu$. Hence, $t_2 > \max(\mu, \mu - t_1) = a$ (say) .

We have,

$$g(t_1, t_2) \propto \left[1 - \exp\left\{-\frac{(t_1 + t_2 - \mu)}{\theta}\right\}\right]^{k-1}$$

$$\cdot \exp\left[\left\{-\frac{(m-k+1)\ (t_1 + t_2 - \mu)}{\theta}\right\}\right] \cdot \exp\left[-\frac{n}{\theta}(t_2 - \mu)\right]$$

$$\propto \sum_{i=0}^{k-1} (-1)^i \binom{k-1}{i} \left[\exp\left\{-(m-k+1+i)\ \frac{t_1}{\theta}\right\}\right.$$

$$\left.\exp\left\{-(m-k+1+n+i)\ \frac{(t_2 - \mu)}{\theta}\right\}\right].$$

Integrating out t_2, the marginal density for t_1 is given by

$$g(t_1) \propto \sum_{i=0}^{k-1} (-1)^i \binom{k-1}{i} \left[\exp\left\{-(m-k+1+i)\frac{t_1}{\theta}\right\}\right.$$

$$\left.\cdot \int_a \exp\left\{-(m-k+1+i+n)\frac{(t_2 - \mu)}{\theta}\right\} d\theta\right].$$

If $y_k < x_1$, $t_1 < 0$ and $a = \mu - t_1$.

If $y_k > x_1$, $t_1 > 0$ and $a = \mu$.

Thus,

$$g(t_1) \propto \sum_{i=0}^{k-1} (-1)^i \binom{k-1}{i} \left[\exp\left\{-(m-k+1+i)\frac{t_1}{\theta}\right\}\right.$$

$$\left.\cdot \int_{\mu - t_1}^{\infty} \exp\left\{-(m-k+1+i+n)\ \frac{t_2 - \mu}{\theta}\right\} d\theta\right],\ t_1 < 0$$

and, $$\propto \sum_{i=0}^{k-1} (-1)^i \binom{k-1}{i} \left[\exp\left\{-(m-k+1+i)\frac{t_1}{\theta}\right\} \right.$$

$$\left. \cdot \int_{\mu}^{\infty} \exp\left\{-(m-k+1+i+n)\ \frac{(t_2-\mu)}{\theta}\right\} d\theta \right], \quad t_1 > 0$$

or, $$g(t_1) \propto \sum_{i=0}^{k-1} (-1)^i \binom{k-1}{i} \cdot$$

$$(m-k+1+i+n)^{-1} \exp\left(\frac{n\, t_1}{\theta}\right), \quad t_1 < 0$$

and,

$$\propto \sum_{i=0}^{k-1} (-1)^i \binom{k-1}{i} (m-k+1+i+n)^{-1} \cdot$$

$$\exp\left[-\left\{(m-k+1+i)\ \frac{t_1}{\theta}\right\}\right], \quad t_1 > 0 .$$

It is well-known that

$\frac{S}{\theta}$ is distributed as gamma with parameter $(n-1)$,

$$S = \sum_{i=1}^{n} (x_i - x_1) .$$

Let $u = \frac{y_k - x_1}{s} = \frac{t_1}{s}$. For fixed s, dt1 = sdu. Using this substitution in $g(t_1)$, we have

$$g(u \mid s) \propto \sum_{i=0}^{k-1} (-1)^i \binom{k-1}{i} \cdot$$

$$(m-k+1+i+n)^{-1} \cdot s \cdot \exp\left(\frac{n\, s\, u}{\theta}\right), \quad u < 0$$

and,

$$\propto \sum_{i=0}^{k-1} (-1)^i \binom{k-1}{i} (m-k+1+i+n)^{-1} \cdot s \cdot$$

$$\exp\left[-\left\{(m-k+1+i)\ \frac{u\,s}{\theta}\right\}\right],\quad u>0.$$

Hence,

$$g(u) \propto \sum_{i=0}^{k-1} (-1)^i \binom{k-1}{i} (m-k+1+i+n)^{-1}$$

$$\int_0^\infty \exp\left(\frac{nsu}{\theta}\right)\exp\left(\frac{-s}{\theta}\right) ds$$

$$\propto \sum_{i=0}^{k-1} (-1)^i \binom{k-1}{i} (m-k+1+i+n)^{-1}\ (1-nu)^{-n},\ u<0$$

and

$$\propto \sum_{i=0}^{k-1} (-1)^i \binom{k-1}{i} (m-k+1+i+n)^{-1}$$

$$\int_0^\infty s^{n-1} \exp\left[-\left\{\frac{(m-k+1+i)\,u+1}{\theta}\right\}\right] ds$$

$$\propto \sum_{i=0}^{k-1} (-1)^i \binom{k-1}{i} (m-k+1+i+n)^{-1}\cdot$$

$$\{m-k+1+i)\,u+1\ \}^{-n},\quad u>0.$$

The predictive distribution of $Y_{(k)}$ may now be written out

$$h\,(y_k) = C \sum_{i=0}^{k-1} (-1)^i \binom{k-1}{i} (m-k+1+i+n)^{-1}\cdot$$

$$\{s+n\,(x_1-y_k)\ \}^{-n},\ y_k < x_1 \equiv h_1\ \text{(say)}$$

and,

$$= C \sum_{i=0}^{k-1} (-1)^i \binom{k-1}{i} (m-k+1+i+n)^{-1}\cdot$$

$$\{s+(m-k+1+i)\ (y_k-x_1)\ \}^{-n},\ y_k \geq x_1 \equiv h_2\quad \text{(say)}$$

where

$$C^{-1} = \int_{0}^{\infty} h(y_k)\, d\, y_k$$

$$= \frac{B(k\,, m-k+1)}{n\,(n-1)\, s^{n-1}} \left(1 - \left(\frac{s}{s+n\,x_1}\right)^{n-1} \left[\frac{B\,(k\,, m-k+n+1)}{B\,(k\,, m-k+1)}\right]\right).$$

The predictive distribution $h(y_k)$ for the two-paratmeter exponential distribution will have exactly the same form, except for the normalizing constnat C where $C^{-1} = \int_{-\infty}^{\infty} h\,(y_k)\, d\, y_k$

$$= \frac{B\,(k\,, m-k+1)}{n\,(n-1)\, s^{n-1}}\,.$$

Predictive Interval (classical)

We now consider $(1-\alpha)$-predictive interval $I = (L\,, U)$ for $y_{(k)}$ where L and U are solutions of

$$\frac{\alpha}{2} = \int_{0}^{L} h_1\, d\, y_k = \int_{U}^{\infty} h_2\, d\, y_k \text{ which leads to}$$

$$L = \bar{x} - \frac{1}{n}\left[\left\{\frac{n(n-1)\,\alpha}{2B\,(k\,, m-n+k+1)}\right\} C^{-1} + (n\,\bar{x})^{-(n-1)}\right]^{-\frac{1}{(n-1)}}$$

and U is the solution of

$$\left\{\frac{\alpha\,(n-1)}{2}\right\} C^{-1} = \sum_{i=0}^{k-1} (-1)^i \binom{k-1}{i}\cdot$$

$$[(m-k+1+i)\ (m-k+1+n+i)]^{-1}\cdot$$

$$\cdot\,[s+(m-k+1+i)\ (U-x_1)]^{-(n-1)}\,.$$

In particular, for $m = k = 1$,

$$L = \bar{x} - \frac{1}{n}\left[\left\{\frac{n\,(n-1)\,(n+1)\,\alpha}{2}\right\} C^{-1} + (n\,\bar{x})^{-(n-1)}\right]^{-\frac{1}{(n-1)}}.$$

$$U = x_1 - s + \left[\left\{\frac{n\,(n-1)\,(n+1)\,\alpha}{2}\right\} C^{-1}\right].$$

For two-parameter exponential case with $-\infty < \mu < \infty$, L and U are solutions of:

$$\frac{\alpha}{2} = \int_{-\infty}^{L} h_1\, d\,y_k = \int_{U}^{\infty} h_2\, d\,y_k \text{ where}$$

U is the solution of

$$\left\{\frac{\alpha\,(n-1)}{2}\right\} C^{-1} = \sum_{i=0}^{k-1} (-1)^i \binom{k-1}{i}$$

$$[(m-k+1+i)\,(m-k+1+n+i)]^{-1}$$

$$\cdot[s+(m-k+1+i)\,(U-x_1)]^{-(n-1)}$$

and,

$$L = \bar{x} - \frac{1}{n}\left[\left\{\frac{n\,(n-1)\,\alpha}{2B\,(k\,,m-k+n+1)}\right\} C^{-1}\right]^{-\frac{1}{(n-1)}},$$

and in this case

$$C^{-1} = \frac{B(k\,,m-k+1)}{n(n-1)\,s^{n-1}}.$$

In particular for $m = k = 1$

$$L = \bar{x} - \frac{s}{n}\left[\frac{\alpha\,(n+1)}{2}\right]^{\left(\frac{1}{n-1}\right)}$$

$$U = x_1 + s\left[\left\{\frac{\alpha\,(n+1)}{2\,n}\right\}^{-\left(\frac{1}{(n-1)}\right)} - 1\right].$$

Numerical example:

We computed 80%, 90%, and 95% prediction intervals for Y_k, using Grubbs' (1971) data such that $n = 19$, $s = 15,869$, $x_{(1)} = 162$, $n\bar{x} = 18{,}947$ for the two models (3.14) and (3.23) and sets of (m, k) (1, 1) and (5, 2).

Table 1

$m = k = 1,$	$-\infty\ \mu < \infty$;	Classical/Bayesian	
$1-\alpha$	L	U	$I(= U - L)$
0.80	193.56	2276.22	2082.66
0.90	162.01	2982.23	2820.22
0.95	129.22	3715.96	3586.74

Table 2

$m = k = 1,$	$\mu \geq 0$;	Classical	
$1-\alpha$	L	U	I
0.80	194.38	2278.28	2083.90
0.90	163.79	2984.37	2820.58
0.95	132.93	3718.18	3585.25

Table 3

$m = 5,\ k = 2,$	$\mu \geq 0$	Classical	
$1-\alpha$	L	U	I
0.80	208.39	793.15	584.76
0.90	178.08	1001.49	823.41
0.95	147.28	1213.26	1065.98

Table 4

$m = 5\ k = 2,$	$\mu \geq 0$	Bayesian	
$1-\alpha$	L	U	I
0.80	222.89	948.60	725.71
0.90	185.56	1156.87	971.31
0.95	158.27	1366.94	1208.67

Table 5

$m = 5,\ k = 2,$	$-\infty < \mu < \infty$;	Classical/Bayesian	
$1-\alpha$	L	U	I
0.80	216.64	948.70	733.06
0.90	177.17	1158.04	980.87
0.95	143.60	1369.79	1226.19

(Sinha, 1995)

Comparing the tables, we observe that for $m = k = 1$ and the given data, while there is not much gain with the left truncated exponential over its two-parameter counterpart, for $m = 5$, $k = 2$, the intervals, classical or Bayesian, are uniformly shorter for $\mu \geq 0$ than when μ is allowed to be negative. Further, Tables 3 and 4 show that for $m = 5$, $k = 2$, $\mu \geq 0$, the classical prediction intervals are uniformly shorter than the corresponding Bayesian results. This is what one would expect since the sample information about the parameter contained in the likelihood would dominate the prior information reflected by the vague prior $g(\mu, \theta) \propto \frac{1}{\theta}$.

3.3 RAYLEIGH DISTRIBUTION

The Rayleigh probability density function of a random variable X is given by

$$f(x \mid \theta) = \frac{x}{\theta^2} \exp\left(-\frac{x^2}{2\theta^2}\right), \quad x, \theta > 0. \tag{3.29}$$

The likelihood function,

$$L(\underline{x} \mid \theta) \propto \frac{1}{\theta^{2n}} \exp\left(-\frac{S^2}{2\,\theta^2}\right), \quad S^2 = \sum_{i=1}^{n} x_i^2.$$

We will consider the predictive distribution of a future observation Y under the ALI prior due to Hartigan (1964). For simplicity we will refer to it as Hartigan prior.

ALI prior for the Rayleigh pdf (3.29) is given by:

$$g(\theta) \propto \frac{1}{\theta^3}. \qquad \text{(Appendix A.6)}$$

The posterior distribution of θ is given by

$$\Pi(\theta \mid \underline{x}) \propto L(\underline{x} \mid \theta)\, g(\theta)$$

$$\propto \frac{1}{\theta^{2n+3}} \exp\left(-\frac{S^2}{2\,\theta^2}\right).$$

The predictive distribution,

$$h(y \mid \underline{x}) \propto \int_0^\infty \Pi(\theta \mid \underline{x})\, f(y \mid \theta)\, d\theta$$

$$\propto \int_0^\infty \frac{y}{\theta^{2n+5}} \exp\left\{-\left(\frac{y^2+S^2}{2\,\theta^2}\right)\right\} d\theta$$

$$= \frac{K\,y}{(y^2+S^2)^{n+2}}, \; y > 0$$

$$K^{-1} = \int_0^\infty \frac{y\,dy}{(y^2+S^2)^{n+2}}$$

$$= \frac{1}{2(n+1)\,S^{2n+2}}.$$

Restoring the normalizing constant

$$h(y \mid \underline{x}) = \frac{2y\,(n+1)S^{2n+2}}{(S^2+y^2)^{n+2}}, \quad y > 0. \tag{3.30}$$

Let the $(1-\alpha)$ — equal tail prediction interval be $[y_1, y_2]$ where

$$\int_0^{y_1} h(y \mid \underline{x})\, dy = \int_{y_2}^{\infty} h(y \mid \underline{x}) dy = \alpha/2.$$

From (3.30)

$$\frac{\alpha}{2} = (n+1)\, S^{2n+2} \int_0^{y_1} \frac{2y\,dy}{(y^2+S^2)^{n+2}}$$

$$= 1 - \left(\frac{S^2}{S^2+y_1^2}\right)^{n+1},$$

or $$y_1 = \left[\left(1-\frac{\alpha}{2}\right)^{\frac{-1}{n+1}} - 1\right]^{\frac{1}{2}} S.$$

Similarly $$y_2 = \left[\left(\frac{\alpha}{2}\right)^{\frac{-1}{n+1}} - 1\right]^{\frac{1}{2}} S$$

and

$$I_H = y_2 - y_1 = \left[\left\{\left(\frac{\alpha}{2}\right)^{\frac{-1}{n+1}} - 1\right\}^{\frac{1}{2}} - \left\{\left(1 - \frac{\alpha}{2}\right)^{\frac{-1}{n+1}} - 1\right\}^{\frac{1}{2}}\right] S \qquad (3.31)$$

where I_H refers to the intervals with respect to Hartigan prior.

Now consider the frequentist or classical approach.

The pdf of $Z = \frac{y^2}{S^2}$ is given by

$$h(z) = \frac{1}{B(1, n)(1+z)^{n+1}}, \quad z > 0.$$

Solving for z_1, z_2, where

$$\frac{\alpha}{2} = \int_0^{z_1} h(z)\, dz = \int_{z_2}^{\infty} h(z)\, dz$$

we have the classical interval for Y

$$I_C = \left[\left\{\left(\frac{\alpha}{2}\right)^{\frac{-1}{n}} - 1\right\}^{\frac{1}{2}} - \left\{\left(1 - \frac{\alpha}{2}\right)^{\frac{-1}{n}} - 1\right\}^{\frac{1}{2}}\right] S. \qquad (3.32)$$

Comparing (3.31) and (3.32) it follows that I_H is uniformly shorter than I_C.

A random sample of size $n = 20$ was generated from the Rayleigh pdf (3.29) which yields $S^2 = 150.1435$. From (3.31) and (3.32) we compute 95% prediction intervals,

$$I_H = \left[\left\{(0.025)^{-\frac{1}{21}} - 1\right\}^{\frac{1}{2}} - \left\{(0.975)^{\frac{-1}{21}} - 1\right\}^{\frac{1}{2}}\right] \sqrt{150.1435} = 4.9440$$

$$I_C = \left[\left\{(0.025)^{\frac{-1}{20}} - 1\right\}^{\frac{1}{2}} - \left\{(0.975)^{\frac{-1}{20}} - 1\right\}^{\frac{1}{2}}\right] \sqrt{150.1435} = 5.0785 > I_H.$$

Random samples of size n = 10, 20, 30 were generated from the pdf (3.25), and I_H and I_C were computed for $\alpha = 0.01, 0.05$ and 0.10.

Prediction intervals for Rayleigh Distribution

$(I_H - I_C)/S$

n \ α	0.01	0.05	0.10
10	0.0481	0.0344	0.0282
20	0.0146	0.0110	0.0092
30	0.0076	0.0058	0.0049

As n increases, the two intervals tend to be identical. (Sinha, 1990)

3.4 WEIBULL DISTRIBUTION

Weibull pdf of a random variable X is given by

$$f(x \mid p, \theta) = \frac{p}{\theta} x^{p-1} \exp\left(-\frac{x^p}{\theta}\right), \quad p, \theta, x > 0. \qquad (3.33)$$

From (2.33), the joint posterior is

$$\Pi(\theta, p \mid \underline{x}) \propto \frac{p^{n-1}}{\theta^{n+1}} \lambda^{p-1} \exp\left(\frac{-\sum x_i^p}{\theta}\right), \lambda = \Pi_{i=1}^n x_i .$$

The predictive distribution of Y is given by

$$h(y \mid \underline{x}) \propto \int_0^\infty \int_0^\infty \frac{p^n}{\theta^{n+2}} \lambda^{p-1} y^{p-1} \exp\left(\frac{\sum_{i=1}^n x_i^p + y^p}{\theta}\right) d\theta \, dp$$

$$= K \int_0^\infty \frac{p^n \lambda^{p-1} y^{p-1} \, dp}{\left(\sum_{i=1}^n x_i^p + y^p\right)^{n+1}},$$

where
$$K^{-1} = \int_0^\infty h(y \mid \underline{x}) \, dy .$$

Restoring the normalizing constant, we obtain

$$h(y \mid \underline{x}) = \frac{n \displaystyle\int_0^\infty \frac{p^n \lambda^{p-1} y^{p-1}\, dp}{\left(\sum_{i=1}^n x_i^p + y^p\right)^{n+1}}}{\displaystyle\int_0^\infty \frac{p^{n-1} \lambda^{p-1}\, dp}{\left(\sum_{i=1}^n x_i^p\right)^n}} \tag{3.34}$$

Computation of prediction interval of Y will require a double-integration programming routine/Mathematica sub-routine.

3.5 NORMAL DISTRIBUTION

$$f(x \mid \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \quad -\infty < \mu < \infty, \ \sigma > 0.$$

Using a vague-prior $g(\mu, \sigma) \propto \frac{1}{\sigma}$, the joint posterior distribution

$$\Pi(\mu, \sigma \mid \underline{x}) \propto L(\underline{x} \mid \mu, \sigma) g(\mu, \sigma)$$

$$\propto \frac{1}{\sigma^{n+1}} \exp\left[-\frac{1}{2\sigma^2}\{A + n(\bar{x}-\mu)^2\}\right] \tag{3.35}$$

where $\quad A = \sum_{i=1}^n (x_i - \bar{x})^2.$

From (3.1), and (3.35) we obtain the predictive distribution of Y given by

$$h(y \mid \underline{x}) \propto \int_0^\infty \int_{-\infty}^\infty \exp\left[-\frac{1}{2\sigma^2}\{A + n(\bar{x}-\mu)^2 + (y-\mu)^2\}\right] \frac{d\mu\, d\sigma}{\sigma^{n+2}}$$

$$\propto \int_0^\infty \int_{-\infty}^\infty \exp\left[-\frac{n+1}{2\sigma^2}\left(\mu - \frac{n\bar{x}+y}{n+1}\right)^2\right]$$

$$\exp\left[-\frac{1}{2\sigma^2}\left\{A+\frac{n(y-\bar{x})^2}{n+1}\right\}\right]\frac{d\mu\, d\sigma}{\sigma^{n+2}}.$$

Using the result $\int_{-\infty}^{\infty} \exp(-kx^2)\,dx = \frac{\sqrt{\pi}}{\sqrt{k}}$ (Appendix A.2)

and integrating with respect to μ

$$\int_{-\infty}^{\infty} \exp\left[-\frac{n+1}{2\sigma^2}\left(\mu-\frac{n\bar{x}+y}{n+1}\right)^2\right]d\mu = \sigma \text{ (constant)}$$

Hence, $h(y \mid \underline{x}) \propto \int_0^{\infty} \exp\left[-\frac{1}{2\sigma^2}\left\{A+\frac{n(y-\bar{x})^2}{n+1}\right\}\right]\frac{d\sigma}{\sigma^{n+1}}$

$$\propto \int_0^{\infty} \frac{\exp\left[-\frac{1}{2\sigma^2}\left\{A+\frac{n(y-\bar{x})^2}{n+1}\right\}d\sigma^2\right]}{(\sigma^2)^{\frac{n}{2}+1}}.$$

Using the result $\int_0^{\infty} \frac{\exp\left(-\frac{a}{x}\right)}{x^{m+1}}\,dx = \frac{\Gamma(m)}{a^m}$ (Appendix A.1).

We have $h(y \mid \underline{x}) \propto \frac{1}{\left[A+\frac{n(y-\bar{x})^2}{n+1}\right]^{\frac{n}{2}}}$

$$= \frac{K}{\left[1+\frac{n(y-\bar{x})^2}{(n+1)A}\right]^{\frac{n}{2}}} \quad -\infty < y < \infty. \tag{3.36}$$

where $K^{-1} = \int_{-\infty}^{\infty} \frac{dy}{\left[1+\frac{n(y-\bar{x})^2}{(n+1)A}\right]^{\frac{n}{2}}}.$

Putting

$$\sqrt{\frac{n}{(n+1)A}}\,(y-\bar{x}) = \sqrt{u}$$

$$dy = \sqrt{\frac{(n+1)A}{n}}\ \frac{1}{2}\, u^{\frac{1}{2}-1}\ du,$$

$$K^{-1} = 2\int_0^\infty \frac{dy}{\left[1+\frac{n(y-\bar{x})^2}{(n+1)A}\right]^{\frac{n}{2}}}$$

$$= \sqrt{\frac{(n+1)A}{n}} \int_0^\infty \frac{u^{\frac{1}{2}-1}\, du}{(1+u)^{\frac{n-1}{2}+\frac{1}{2}}}$$

$$= \sqrt{\frac{(n+1)A}{n}} \cdot B\left(\frac{1}{2}, \frac{n-1}{2}\right),$$ using the beta-integral

$$\int_0^\infty \frac{x^a\, dx}{(1+x)^{a+b}} = B(a, b) \qquad \text{(Appendix A.7)}$$

Substituting K in (3.36), we get

$$h(y \mid \underline{x}) = \frac{\sqrt{n}}{\sqrt{(n+1)A}\ B\left(\frac{1}{2}, \frac{n-2}{2}\right)} \cdot \frac{1}{\left[1+\frac{n(y-\bar{x})^2}{(n+1)A}\right]^{\frac{n}{2}}}.$$

Putting $$\frac{\sqrt{n}\,(y-\bar{x})}{\sqrt{(n+1)A}} = \frac{w}{\sqrt{n-1}}.$$

$$dy = \sqrt{\frac{(n+1)A}{n(n-1)}}\, dw$$

we have

$$\rho(w) = \frac{1}{\sqrt{n-1}\ B\left(\frac{1}{2}, \frac{n-1}{2}\right)} \cdot \frac{1}{\left(1+\frac{w^2}{n-1}\right)^{\frac{n}{2}}} \quad -\infty < w < \infty \qquad (3.37)$$

which is Student's-t distribution with $(n-1)$ of freedom.

Thus, it follows from (3.37) that $\sqrt{\frac{n(n-1)}{(n+1)A}}\ (y-\bar{x})$ is distributed as Student's-t with $\nu = (n-1)$ degrees of freedom. Hence,

$$1-\alpha = P\left[-t\left(\frac{\alpha}{2}, \nu\right) < t < t\left(\frac{\alpha}{2}, \nu\right)\right]$$

$$= P\left[-t\left(\frac{\alpha}{2}, \nu\right) < \sqrt{\frac{n(n-1)}{(n+1)A}}\,(y-\bar{x}) < t\left(\frac{\alpha}{2}, \nu\right)\right]$$

$$= P\left[\bar{x} - t\left(\frac{\alpha}{2}, \nu\right)\sqrt{\frac{(n+1)A}{n(n-1)}} < Y < \bar{x}\right.$$

$$\left.+\, t\left(\frac{\alpha}{2}, \nu\right)\sqrt{\frac{(n+1)A}{n(n-1)}}\right].$$

Then, $(1-\alpha)$ predictive limits of Y are

$$\bar{x} \pm t\left(\frac{\alpha}{2}, \nu\right)\sqrt{\frac{(n+1)A}{n(n-1)}}$$

$$= \bar{x} \pm t\left(\frac{\alpha}{2}, \nu\right) s\sqrt{\left(1 + \frac{1}{n}\right)},$$

where, $$s^2 = \frac{A}{(n-1)} = \frac{\sum_{i=1}^{n}(x_i - x)^2}{n-1},$$

and $t(\alpha, \nu)$ represents 100 α % point of student's-t distribution with ν degrees of freedom.

It is interesting to consider the corresponding classical counterpart and compare the two intervals.

Let $\bar{x}$ be the mean of random sample $(x_1, x_2, \cdots, x_n)$ from $N(\mu, \sigma^2)$ and Y be the future observation.

$$\text{Var}(Y - \bar{X}) = \sigma^2\left(1 + \frac{1}{n}\right)$$

$$\frac{(Y-\bar{X})^2}{\sigma^2\left(1 + \frac{1}{n}\right)} \sim \chi^2 \text{ with 1 degree of freedom}$$

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2 \text{ with } (n-1) \text{ degrees of freedom}$$

since $(Y-\bar{X})^2$ and S^2 are independently distributed, the ratio

$$\frac{(Y-\bar{X})^2}{\frac{(n+1)(n-1)S^2}{n}} \sim \frac{t^2}{n-1} \quad \text{and}$$

$$\frac{(Y-\bar{x})}{\sqrt{\frac{(n+1)S^2}{n}}} \sim t \text{ with } (n-1) \text{ degrees of freedom.}$$

Thus, the classical $(1-\alpha)$-predictive limits are $\bar{x} \pm t\left(\frac{\alpha}{2}, n-1\right)$. $s\sqrt{\frac{n+1}{n}}$ which are the same as their Bayesian counterparts.

Predictive distribution/prediction interval of $Y \sim$ lognormal (μ, σ^2) is left as an exercise for the readers.

3.6 PREDICTIVE DISTRIBUTION AND RELIABILITY ESTIMATION

Let the random variable X represent the life of an item or a component and let $f(x \mid \theta)$ be the probability density function of X.

The reliability function at any time t is defined by

$$R_t = P(X \geq t)$$

$$= \int_t^{\infty} f(x \mid \theta)\, dx.$$

Under squared-error loss function Bayes estimate of R_t is given by

$$R_t^* = E_\theta(R_t \mid \underline{x})$$

$$= E_\theta[P(Y \geq t \mid \underline{x})]$$

$$= E_\theta\left[\int_t^{\infty} f(y \mid \theta)\, dy \mid \underline{x}\right]$$

$$= \int_{\Omega} \pi(\theta \mid \underline{x}) \left[\int_{t}^{\infty} f(y \mid \theta)\, dy \right] d\theta$$

$$= \int_{t}^{\infty} \left[\int_{\Omega} f(y \mid \theta)\, \pi(\theta \mid \underline{x})\, d\theta \right] dy$$

and from (3.1)

$$R_t^{*} = \int_{t}^{\infty} h(y \mid \underline{x})\, dy \tag{3.38}$$

= Reliability function of the predictive distribution.

Thus, under squared-error loss function, Bayes estimator of the reliability function of a distribution is the reliability funciton of the predictive distribution.

We will now derive reliability functions for the following distributions using the definition (3.38).

Exponential Distributions

(a) Consider the one-parameter exponential pdf (3.3).

From (3.5) and (3.42)

$$R_t^{*} = n S^n \int_{t}^{\infty} \frac{dy}{(S+y)^{n+1}}$$

$$= n S^n \int_{s+t}^{t} \frac{d\mu}{u^{n+1}}$$

$$= \left(\frac{S}{S+t} \right)^n$$

$$= \frac{1}{\left(1 + \frac{t}{S}\right)^n}, \quad t > 0. \tag{3.39}$$

Note that whereas the uniformly-minimum variance-unbiased estimator of R_t for this distribution is

$$= \int_{\Omega} \Pi(\theta \mid \underline{x}) \left[\int_{t}^{\infty} f(y \mid \theta)\, dy \right] d\theta$$

$$= \int_{t}^{\infty} \left[\int_{\Omega} f(y \mid \theta)\, \Pi(\theta \mid \underline{x})\, d\theta \right] dy$$

and from (3.1)

$$R_t^* = \int_{t}^{\infty} h(y \mid \underline{x})\, dy \tag{3.38}$$

= Reliability function of the predictive distribution.

Thus, under squared-error loss function, Bayes estimator of the reliability function of a distribution is the reliability funciton of the predictive distribution.

We will now derive reliability functions for the following distributions using the definition (3.38).

Exponential Distributions

(a) Consider the one-parameter exponential pdf (3.3).

From (3.5) and (3.42)

$$R_t^* = n\, S^n \int_{t}^{\infty} \frac{dy}{(S+y)^{n+1}}$$

$$= n\, S^n \int_{s+t}^{t} \frac{d\mu}{u^{n+1}}$$

$$= \left(\frac{S}{S+t} \right)^n$$

$$= \frac{1}{\left(1 + \frac{t}{S}\right)^n}, \quad t > 0. \tag{3.39}$$

Note that whereas the uniformly-minimum variance-unbiased estimator of R_t for this distribution is

$$R_t = \left(1 - \frac{t}{S}\right)^{n-1} \qquad t < S$$

and, $\quad = 0 \quad t \geq S$ (Basu, 1964)

Bayes estimator R_t^* given by (3.39) is always positive.

(b) For the left-truncated exponential model (3.23):

(i) $\mu \geq 0, \quad t \leq x_1$.
From (3.25) and (3.26) we have

$$R_t^* = \int_t^{x_1} h_1 \, d y_k + \int_{x_1}^{\infty} h_2 \, d y_k$$

$$= C \sum_{i=0}^{k-1} (-1)^i \binom{k-1}{i} (m-k+1+i)^{-1} \cdot$$

$$\left[\int_t^{x_1} \{s + n(x_1 - y_k)\}^{-n} d y_k \right.$$

$$+ \int_{x_1}^{\infty} \{s + (m-k+1+i)(y_k - x_1)\}^{-n} d y_k$$

$$\left. - \int_t^{\infty} \{s + n x_1 + y_k (m-k+1+i)^{-n}\} d y_k \right]$$

$$= \frac{C}{n(n-1)} s^{n-1} \left[B(k, m-k+1) - B(k, m-k+n+1) \cdot \right.$$

$$\left\{1 + \frac{n(x_1 - t)}{s}\right\}^{-(n-1)} - n \sum_{i=0}^{k-1} (-1)^i \binom{k-1}{i} \cdot$$

$$(m-k+1+n+i)^{-1} (m-k+1+i)^{-1} \cdot$$

$$\left. \left\{1 + \frac{n x_1 + t(m-k+1+i)}{s}\right\}^{-(n-1)} \right]$$

where C is defined in (3.24).

(ii) $\mu \geq 0, \; t \geq x_1$.

From (3.25), after some algebra we obtain

$$R_t^* = \int_t^\infty h_2 \, d y_k = \frac{C}{(n-1)\, s^{n-1}} \left[\sum_{i=0}^{k-1} (-1)^i \binom{k-1}{i} \cdot \right.$$

$$((m-k+1+i)\,(m-k+1+n+i))^{-1} \cdot$$

$$\left\{ \left(1 + \frac{(m-k+1+i)\,(t-x_1)}{s} \right)^{-(n-1)} \right.$$

$$\left. \left. - \left(1 + \frac{n\,x_1 + t\,(m-k+1+i)}{s} \right)^{-(n-1)} \right\} \right] .$$

For $m = k = 1$, we have

$$R_t^* = \frac{C}{(n+1)\, n\, (n-1)\, S^{n-1}} \cdot$$

$$\left[(n+1) - \left\{ \left(1 + \frac{n\,(x_1 - t)}{s} \right)^{-(n-1)} \right. \right.$$

$$\left. \left. + n \left(1 + \frac{nx_1 + t}{s} \right)^{-(n-1)} \right\} \right], \quad t \le x_1 .$$

and

$$= \frac{C}{n\,(n-1)\,(n+1)\, s^{n-1}} \cdot$$

$$\left[\left(1 + \frac{t - x_1}{s} \right)^{-(n-1)} - \left(1 + \frac{nx_1 + t}{s} \right)^{-(n-1)} \right], \quad t \ge x_1$$

$R_1^* \to 1$ as $t \to 0$

and $\to 0$ and $t \to \infty$. (Sinha, 1986)

(c) For the two-parameter exponential model (3.14):

(iii) $-\infty < \mu < \infty$, $t \le x_1$.

From (3.18) and (3.19), after some algebra, we obtain

From (3.25), after some algebra we obtain

$$R_t^* = \int_t^{\infty} h_2 \, d\,y_k = \frac{C}{(n-1)\, s^{n-1}} \left[\sum_{i=0}^{k-1} (-1)^i \binom{k-1}{i} \cdot \right.$$

$$((m-k+1+i)\;(m-k+1+n+i))^{-1} \cdot$$

$$\left\{ \left(1 + \frac{(m-k+1+i)\,(t-x_1)}{s}\right)^{-(n-1)} \right.$$

$$\left. \left. - \left(1 + \frac{n\,x_1 + t\,(m-k+1+i)}{s}\right)^{-(n-1)} \right\} \right] .$$

For $m = k = 1$, we have

$$R_t^* = \frac{C}{(n+1)\, n\, (n-1)\, S^{n-1}} \cdot$$

$$\left[(n+1) - \left\{ \left(1 + \frac{n\,(x_1 - t)}{s}\right)^{-(n-1)} \right. \right.$$

$$\left. \left. + \, n\left(1 + \frac{nx_1 + t}{s}\right)^{-(n-1)} \right\} \right], \quad t \le x_1$$

and

$$= \frac{C}{n\,(n-1)\,(n+1)\, s^{n-1}}$$

$$\left[\left[\left(1 + \frac{t - x_1}{s}\right)\right]^{-(n-1)} - \left(1 + \frac{nx_1 + t}{s}\right)^{-(n-1)} \right], \quad t \ge x_1$$

$R_1^* \to 1$ as $t \to 0$ (Sinha, 1986)

and $\to 0$ and $t \to \infty$..

(c) For the two-parameter exponential model (3.14):

(iii) $-\infty < \mu < \infty, \; t \le x_1$.

From (3.18) and (3.19), after some algebra, we obtain

$$R_t^* = A \sum_{i=0}^{k-1} (-1)^i \binom{k-1}{i} (m-k+1+n+i)^{-1} \cdot$$

$$\left[\int_t^{x_1} \{s + n(x_1 - y_k)\}^{-n} \, dy_k \right.$$

$$\left. + \int_{x_1}^{\infty} \{s + (m-k+1+i)(y_k - x_1)\}^{-n} \, dy_k \right]$$

$$= 1 - \frac{(B(k, m-k+n+1)}{B(k, m-k+1)} \left\{ 1 + \frac{n(x_1 - t)}{s} \right\}^{-(n-1)},$$

$$A^{-1} = \frac{B(k, m-k+1)}{n(n-1)_s^{n-1}}$$

(iv) $-\infty < \mu < \infty,\ t \geq x_1$.

From (3.18), we obtain

$$R_t^* = A \sum_{i=0}^{k-1} (-1)^i \binom{k-1}{i} (m-k+1+i+n)^{-1} \cdot$$

$$\int_t^{\infty} \{s + (m-k+1+i)^{-1} (y_k - x_1)\}^{-n} \, dy_k$$

$$= \frac{n}{B(k, m-k+1)} \sum_{i=0}^{k-1} (-1)^i \cdot$$

$$\binom{k-1}{i} \{(m-k+1+i)(m-k+1+i+n)\}^{-1} \cdot$$

$$\left[1 + \frac{(m-k+1+i)(t-x_1)}{s} \right]^{-(n-1)}.$$

From $m = k = 1$,

$$R_t^* = 1 - \frac{1}{n+1} \left\{ 1 + \frac{n(x_1 - t)}{s} \right\}^{-(n-1)} \qquad t \leq x_1$$

and,
$$= \frac{n}{n+1} \left\{ 1 + \frac{t - x_1}{s} \right\} \qquad t \geq x_1 .$$

Since μ is allowed to assume –ve values, we observe $R_t^* \rightarrow 1$ as $t \rightarrow -\infty$ (only of academic interest) and $\rightarrow 0$ as $t \rightarrow \infty$.

Example 3.1

Using Grubb's (1971) data, we compute R_t^* for $(m,k) = \{(1,1) ; (2,5)\}$ at arbitrary $t = 150 < x_1$ and $200 > x_1$.

Table 6: R_t^* $(m = k = 1)$

μ \ t	150	200
$\mu \geq 0$	0.9151	0.9153
$-\infty < \mu < \infty$	0.9613	0.9100

Table 7: R_t^* $(m = 5, k = 2)$

μ \ t	150	200
$\mu \geq 0$	0.9470	0.9987
$-\infty < \mu < \infty$	0.9720	0.9232

For the two sets of (m,k) it appears from Tables 6 and 7 that the reliability estimates are more sensitive for $\mu \geq 0$ than the corresponding situation $-\infty < \mu < \infty$.

The location parameter μ is a warranty parameter and hence, is essentially non-negative. The two-parameter exponential model where μ is allowed to assume –ve values, is unacceptable in life testing problems. Further, under this model $R_t^* \rightarrow 1$ as $t \rightarrow -\infty$ has no sensible interpretation except for some academic interest. The predictive limits are obtained in simple closed forms. Using Grubbs (1971) data we observe that the prediction intervals, classical or Bayesian, are uniformly shorter for $\mu \geq 0$ than when μ varies over the entire real time. Thus, at least for Life testing/reliability problems, the left-truncated exponential is more meaningful than the two-parameter model.

Normal Distribution

Let $\underline{x} = (x_1, x_, \cdots, x_n)$ be a random sample from a normal distribution $N(\mu, \sigma^2)$ and let y be a future observation from the same distribution.

We have shown in (3.37) that

$$W = \sqrt{\frac{n(n-1)}{(n+1)S}}\ (Y-\bar{x})$$

is distributed as Student's-t with $(n-1)$ degrees of freedom.

$$R_t^* = \int_t^{\infty} h(y \mid \underline{x})\ dy = \int_{\sqrt{\frac{n(n-1)}{(n+1)A}}(t-\bar{x})} g(u \mid \underline{x})\ du$$

$$= 1 - F_U\left\{\sqrt{\frac{n(n-1)}{(n+1)A}}\ (t-\bar{x})\right\} \qquad (3.40)$$

Example 3.2:

The following is a random sample of size $n = 10$ from a $N(\mu, \sigma^2)$ with $\mu = 310$, $s = 180$. $x_i = 392, 252, 176, 232, 146, 100, 244, 700, 476, 364$.

Obtain Bayes estimate of R_t at $t = 300$.

We have $\bar{x} = 308.2$, $A = 292019.6$.

From (3.40) we have

$$R_{300}^* = 1 - F_u\left\{\sqrt{\frac{10\,(9)}{11\,(292019.6)}}\ (300 - 308.2)\right\} = 0.5168\,.$$

Weibull Distribution

Consider the Weibull pdf (3.33)

In (3.34) we have obtained

$$h(y \mid \underline{x}) = \frac{n\displaystyle\int_0^{\infty} \frac{p^n \lambda^{p-1} y^{p-1}\,dp}{\left(\displaystyle\sum_{i=1}^{n} x_i^p + y^p\right)^{n+1}}}{\displaystyle\int_0^{\infty} \frac{p^{n-1}\lambda^{p-1}}{\left(\displaystyle\sum_{i=1}^{n} x_i^p\right)^{n}}\,dp}$$

Hence,

$$R_t^* = \frac{n \int_0^\infty p^n \lambda^{p-1} \left[\int_t^\infty \frac{y^{p-1}\, dy}{\left(\sum_{i=1}^n x_i^p + y^p \right)^{(n+1)}} \right] dp}{\int_0^\infty \frac{p^{n-1} \lambda^{p-1}\, dp}{\left(\sum_{i=1}^n x_i^p \right)^n}}$$

$$= \frac{\int_0^\infty \frac{p^{n-1} \lambda^{p-1}\, dp}{\left(\sum_{i=1}^n x_i^p + t^p \right)^n}}{\int_0^\infty \frac{p^{n-1} \lambda^{p-1}\, dp}{\left(\sum_{i=1}^n x_i^p \right)^n}}$$

which may be evaluated using the numerical integration routine Mathematica.

Rayleigh Distribution

The Rayleigh-random variable X has the pdf

$$f(x \mid \theta) = \frac{x}{\theta^2} \exp\left(- \frac{x^2}{2\theta^2} \right), \quad x,\ \theta > 0\,. \tag{3.41}$$

From (3.30) we have

$$h(y \mid \underline{x}) = \frac{2y\,(n+1)\, S^{2n+2}}{(S^2 + y^2)^{(n+2)}}, \quad y > 0\,.$$

Hence,

$$R_t^* = 2(n+1)\; S^{2n+2} \int_t^\infty \frac{y\, dy}{(S^2 + y^2)^{(n+2)}}$$

$$= \frac{S^{2n+2}}{(S^2 + t^2)^{n+1}}$$

$$= \frac{1}{\left(1 + \frac{t^2}{S^2}\right)^{n+1}}, \quad t > 0 .$$

EXERCISES

1. Let $\underline{x} = (x_1, x_2, \cdots, x_n)$, $n = 15$ be a random sample from the exponential pdf (3.3).

   (i) Plot the predictive distribution (3.5) and construct the 90% prediction interval for a future observation Y.

   (ii) Compute R_t^* at $t = 2.0$.

   Data:

8.8288	11.0500	4.1630	22.8332	0.8989
0.0285	20.4813	4.9149	5.9788	7.1655
2.0921	0.6309	0.8258	0.9521	7.5306

2. Use the data in Exercise 1 above, and construct the 90% prediction interval of Y with respect to the prior $g(\theta) \propto \frac{1}{\theta^2}$, $\theta > 0$.

3. Derive the reliability function R_t^* for the two-parameter exponential distribution (3.14) under the prior $g(\mu, \theta) \propto \frac{1}{\theta}$.

4. Compute the 90% prediction intervals for $Y_{(k)}$ from the exponential pdf (3.14) for $\alpha = 0.05, 0.10, 0.20$ given $g(\mu, \theta) \propto \frac{1}{\theta}$, $\underline{x} = (x_1, x_2, \cdots, x_n)$, $n = 20$, $k = 3$, $m = 10$, $x_{(1)} = 5.0077$ and $S = 2327.8840$.

5. Construct the same set of prediction intervals using the data in Exercise 4, assuming the underlying distribution is left-

$$= \frac{S^{2n+2}}{(S^2 + t^2)^{n+1}}$$

$$= \frac{1}{\left(1 + \frac{t^2}{S^2}\right)^{n+1}}, \quad t > 0.$$

EXERCISES

1. Let $\underline{x} = (x_1, x_2, \cdots, x_n)$, $n = 15$ be a random sample from the exponential pdf (3.3).

   (i) Plot the predictive distribution (3.5) and construct the 90% prediction interval for a future observation Y.

   (ii) Compute R_t^* at $t = 2.0$.

   Data:

8.8288	11.0500	4.1630	22.8332	0.8989
0.0285	20.4813	4.9149	5.9788	7.1655
2.0921	0.6309	0.8258	0.9521	7.5306

2. Use the data in Exercise 1 above, and construct the 90% prediction interval of Y with respect to the prior $g(\theta) \propto \frac{1}{\theta^2}$, $\theta > 0$.

3. Derive the reliability function R_t^* for the two-parameter exponential distribution (3.14) under the prior $g(\mu, \theta) \propto \frac{1}{\theta}$.

4. Compute the 90% prediction intervals for $Y_{(k)}$ from the exponential pdf (3.14) for $\alpha = 0.05, 0.10, 0.20$ given $g(\mu, \theta) \propto \frac{1}{\theta}$, $\underline{x} = (x_1, x_2, \cdots, x_n)$, $n = 20$, $k = 3$, $m = 10$, $x_{(1)} = 5.0077$ and $S = 2327.8840$.

5. Construct the same set of prediction intervals using the data in Exercise 4, assuming the underlying distribution is left-

truncated exponential given by (3.23). How do the two sets of intervals compare?

6. Use the data { S, n, $x_{(1)}$ } in Exercise 4 and compute the reliability functions R_t^* at $t = 4$ and $t = 6$, assuming the underlying distribution is

   (i) two-parameter exponential

   (ii) left-truncated exponential.

7. Let $\underline{x} = (x_1, x_2, \cdots, x_n)$, $n = 30$ be a random sample from the Rayleigh pdf (3.29)

   (i) Plot the predictive density $h(y \mid \underline{x})$.

   (ii) Construct 95% prediction interval of a future observation Y with respect to priors,

   $$g(\theta) \propto \frac{1}{\theta}, \text{ and}$$

   $$g(\theta) \propto \frac{1}{\theta^2}.$$

   (iii) How do the two sets of intervals compare?

   Data: { x_i }.

1.8637	9.7849	10.8185	6.3818	7.6122
4.9225	5.5530	6.0044	3.7891	4.5566
2.5716	3.0413	3.4862	6.6899	7.6331
9.7437	9.8272	12.7716	3.9870	4.7411
5.0859	5.8119	6.1669	6.8934	9.6217
2.8264	3.4174	3.6742	4.5439	15.9188

8. Let $\underline{x} = (x_1, x_2, \cdots, x_n)$ be a random sample from the Rayleigh pdf (3.29). Show that the $(1 - \alpha)$–prediction interval for a future observation Y is uniformly shorter under Hartigan (1964) prior $g(\theta) \propto \frac{1}{\theta^3}$ than the corresponding interval using Jeffreys' (1961) prior $g(\theta) \propto \frac{1}{\theta}$.

9. Let $(x_1, x_2, \cdots, x_n)$, $n = 20$, be a random sample from the lognormal pdf (2.9).

Data:

3.3505	1.2486	1.7446	0.0235
15.8709	24.5633	5.8319	7.5749
0.6340	0.4505	0.8262	9.7556
0.0714	3.2337	0.2281	24.8038
5.7660	1.0266	84.8685	0.0879

Derive and plot the predictive distribution of a future observation Y under the joint prior $g(\mu, \sigma) \propto \frac{1}{\sigma}$.

10. Use the data in Exercise 9 above to construct a 95% prediction interval for Y.

CHAPTER

4 Bayesian Interval Estimation

4.1 CREDIBLE AND HIGHEST POSTERIOR DENSITY (HPD) INTERVALS

Bayesians interpret θ as a random variable and hence, in Bayesian context, it is perfectly logical to ask the following question:

What is the probability that θ lies within a specified interval $[\theta_1, \theta_2]$, $\theta_1 < \theta_2$? Edwards, Lindman and Savage (1963) call this interval based on the posterior distribution of θ, a "Credible interval", an interval that contains a certain fraction of one's degree of belief such that

$$\begin{aligned} 1-\alpha &= P(\theta_1 < \theta < \theta_2) \\ &= \int_{\theta_1}^{\theta_2} \Pi(\theta \mid \underline{x})\, d\theta \end{aligned} \tag{4.1}$$

where $\Pi(\theta \mid x)$ is the posterior distribution of θ. Such an interval $[\theta_1, \theta_2]$ is known as the $(1-\alpha)$-credible interval of θ. An equal-tail $(1-\alpha)$-credible interval $[\theta_1, \theta_2]$ is given by:

$$\frac{\alpha}{2} = \int_{-\infty}^{\theta_1} \Pi(\theta \mid \underline{x})\, d\theta = \int_{\theta_2}^{\infty} \Pi(\theta \mid \underline{x})\, d\theta.$$

To obtain the shortest $(1-\alpha)$-credible interval, one has to minimize $l = \theta_2 - \theta_1$ subject to the condition (4.1) which requires

$$\Pi(\theta_1 \mid \underline{x}) = \Pi(\theta_2 \mid \underline{x}). \tag{4.2}$$

The interval I which simultaneously satisfies (4.1) and (4.2) is called the shortest $(1-\alpha)$-credible interval.

An interval I which satisfies the following conditions simultaneously:

(i) that the interval be shortest possible, and

(ii) $\Pi(\theta \mid \underline{x}),\ \theta \in I \rangle \Pi(\theta \mid \underline{x}),\ \theta \notin I$,

i.e., the posterior density at every point inside be $\rangle$ the posterior density for every point outside it (which of course, implies that the interval includes more probable values of θ and excludes less probable values of the parameter), is called the Highest Posterior Density or HPD Interval (Box and Tiao, 1973).

If the posterior distribution is unimodal but not necessarily symmetric, the shortest credible and HPD intervals are the one and the same (Evans, 1976).

Bayesian confidence interval (Lindley, 1965) differs from the classical or frequentist confidence interval due to the difference in the interpretation of the parameter θ by the two schools. To construct a $(1-\alpha)$-confidence interval for θ we seek a function $g(\underline{x}, \theta)$ of the parameter θ and the observations $\underline{x}$ such that the distribution of $g(\underline{x}, \theta)$ does not depend on any unknown parameter. From the distributional property of $g(\underline{x}, \theta)$ we choose two numbers A and B such that:

$$1-\alpha = P[A < g(\underline{x}, \theta) < B]$$

and work our way through till we end up with the statement,

$$1-\alpha = P\,[L(x) < \theta < U(x)].$$

It will be grossly incorrect to interpret the above relationship as probability is $(1-\alpha)$ that θ lies between $L(x)$ and $U(x)$ since θ is a constant, and it is meaningless to talk about the probability of a constant. What it means is that we are *100 $(1-\alpha)$ % confident* that in repeated sampling from the assumed distribution, a random interval such that

$[L(x),\ U(x)]$ covers the parameter θ. Note that $[L(x),\ U(x)]$ is a realization of a family of random intervals $[L_i(x),\ U_i(x)]$ that we could construct with repeated sampling. Thus, for the classical confidence interval, θ is a constant and the end-points are random variables, and in Bayesian credible interval, θ is a random variable and the end-points are fixed.

4.2 EXPONENTIAL DISTRIBUTIONS

Consider the one-parameter exponential probability density function (pdf),

$$f(x \mid \theta) = \frac{1}{\theta} \exp\left(-\frac{x}{\theta}\right), \quad x,\ \theta > 0. \tag{4.3}$$

From (2.21)′ we have the posterior distribution

$$\Pi(\theta \mid \underline{x}) = \frac{S^n}{\theta^{n+1}\,\Gamma(n)} \exp\left(-\frac{S}{\theta}\right) \quad S = \sum_{i=1}^{n} x_i$$

It follows that $\frac{2S}{\theta} \sim \chi^2(2n)$

$$\begin{aligned} 1-\alpha &= P\left[\chi_1^2 < \chi^2 < \chi_2^2\right] \\ &= P\left[\chi_1^2 < \frac{2S}{\theta} < \chi_2^2\right] \\ &= P\left[\frac{2S}{\chi_2^2} < \theta < \frac{2S}{\chi_1^2}\right], \end{aligned}$$

where $\chi_1^2 = \chi^2\left(1-\frac{\alpha}{2},\ 2n\right)$, $\chi_2^2 = \chi^2\left(\frac{\alpha}{2},\ 2n\right)$, and

$\chi^2(p,\ \nu)$ = upper $100p$ % point of a χ^2-distribution with ν degrees of freedom.

Let $\left[C_L^{(\theta)},\ C_U^{(\theta)}\right]$ and $\left[H_L^{(\theta)},\ H_U^{(\theta)}\right]$ respectively represent the lower and upper credible and HPD interval of a parameter θ.

Thus, the $100(1-\alpha)$ % equal-tail credible interval for the exponential parameter θ is given by

$$\left[C_L^{(\theta)},\ C_U^{(\theta)}\right] = \left[\frac{2S}{\chi^2\left(\frac{\alpha}{2},\ 2n\right)},\ \frac{2S}{\chi^2\left(1-\frac{\alpha}{2},\ 2n\right)}\right] \tag{4.4}$$

which is the same as the classical confidence interval.

The posterior distribution of θ is unimodal and hence, as defined in (4.1) and (4.2), the HPD interval of θ should simultaneously satisfy,

$$1-\alpha = \int_{H_L^{(\theta)}}^{H_U^{(\theta)}} \pi\ (\theta \mid \underline{x})\ d\theta,$$

and $$\Pi\left\{H_L^{(\theta)} \mid \underline{x}\right\} = \Pi\left\{H_U^{(\theta)} \mid \underline{x}\right\}$$

which implies that we have

$$1-\alpha = P\left[H_L^{(\theta)} < \theta < H_U^{(\theta)}\right]$$

$$= P\left[\frac{2S}{H_U^{(\theta)}} < \chi^2\ (2n) < \frac{2S}{H_L^{(\theta)}}\right] \tag{4.5}$$

and, $$\exp\left[-S\left\{\frac{1}{H_L^{(\theta)}} - \frac{1}{H_U^{(\theta)}}\right\}\right] = \left[\frac{H_L^{(\theta)}}{H_U^{(\theta)}}\right]^{(n+1)} \tag{4.6}$$

simultaneously satisfied.

Example 4.1

Suppose we have a random sample $\{x_i\}$ $i = 1, 2, \ldots 15$ from the one-parameter exponential distribution (4.3). We will obtain 95% credible and HPD intervals for θ.

Data : $\{x_i\}$

0.5142	7.8446	8.3758	10.5329	13.9725
3.9769	3.9779	4.3163	4.4317	4.7371
0.6454	0.8451	1.0566	1.2681	15.3301

$$S = \sum_{i=1}^{15} x_i = 81.8252.$$

(i) From (4.4) we have 95% credible interval

$$\left[C_L^{(\theta)},\ C_U^{(\theta)}\right] = \left[\frac{163.6504}{\chi^2\ (0.025, 30)},\ \frac{163.6504}{\chi^2\ (0.975, 30)}\right]$$

$$= [3.4835,\ 9.7470]$$

$$C_U^{(\theta)} - C_L^{(\theta)} = 6.2635$$

From (4.5) and (4.6)

$$\left[H_L^{(\theta)},\ H_U^{(\theta)}\right] = [3.2450,\ 9.1770]$$

$$H_U^{(\theta)} - H_L^{(\theta)} = 5.9320 < C_U^{(\theta)} - C_L^{(\theta)} \text{ as expected.}$$

We now consider the two-parameter left-truncated exponential pdf,

$$f(x \mid \mu, \theta) = \frac{1}{\theta} \exp\left(-\frac{x-\mu}{\theta}\right), \quad 0 < \mu < x_{(1)}, \quad \theta > 0.$$

From (2.28) and (2.28)′ we have the marginal posteriors

$$\Pi_1\ (\mu \mid \underline{x}) = \frac{K}{\{S + n\ (x_{(1)} - \mu\}^n}, \quad 0 < \mu < x_{(1)}$$

and,

$$\Pi_2\ (\theta \mid \underline{x}) = \frac{K}{\Gamma\ (n+1)\ \theta^n}\left[\exp\left(-\frac{S}{\theta}\right) - \exp\left\{-\frac{S + n\, x_{(1)}}{\theta}\right\}\right] \quad \theta > 0$$

where $$K = n\ (n-1)\ S^{n-1}\left[1 - \left\{\frac{S}{S + nx_{(1)}}\right\}^{n-1}\right]^{-1} \tag{4.7}$$

$(1-\alpha)$-credible interval of μ is given by

$\left[C_L^{(\mu)},\ C_U^{(\mu)}\right] \equiv [\mu_1,\ \mu_2]$ where μ_1, μ_2 are solutions of

$$\frac{\alpha}{2} = \int_0^{x_{(1)}} \Pi_1\ (\mu \mid \underline{x})\ d\mu = \int_{\mu_2}^{x_{(1)}} \Pi_1\ (\mu \mid \underline{x})\ d\mu.$$

We have

$$\frac{\alpha}{2} = K \int_0^{\mu_1} \frac{d\mu}{\{S + n\ (x_{(1)} - \mu)\}^n}$$

$$= \frac{K}{n(n-1)} \left[\frac{1}{\{S + n(x_{(1)} - \mu_1)\}^{n-1}} - \frac{1}{\{S + nx_{(1)}\}^{n-1}} \right]$$

$$= \frac{K}{n^n (n-1)} \left[\frac{1}{(\bar{x} - \mu_1)^{n-1}} - \frac{1}{(\bar{x})^{n-1}} \right]$$

which leads to

$$\mu_1 = \bar{x} - \left[\frac{n^n (n-1)\alpha}{2K} + \frac{1}{(\bar{x})^{n-1}} \right]^{\frac{-1}{(n-1)}} \tag{4.8}$$

and, similarly

$$\mu_2 = \bar{x} - \left[\left(\frac{n}{S}\right)^{n-1} - \frac{n^n (n-1)\alpha}{2K} \right]^{\frac{-1}{(n-1)}} \tag{4.9}$$

It follows from (2.27) that $\Pi_1(\mu \mid \underline{x})$ is monotonic in μ, $0 < \mu < x_{(1)}$ and the HPD interval for μ does not exist.

Similarly, the $(1-\alpha)$-credible interval for θ is given by $\left[C_L^{(\theta)}, C_U^{(\theta)}\right] \equiv [\theta_1, \theta_2]$, where θ_1, θ_2 are the solutions of the equations

$$\frac{\alpha}{2} = \int_0^{\theta_1} \Pi_2(\theta \mid \underline{x})\, d\theta = \int_{\theta_2}^{\infty} \Pi_2(\theta \mid \underline{x})\, d\theta.$$

We have

$$\frac{\alpha}{2} = \int_0^{\theta_1} \Pi_2(\theta \mid \underline{x})\, d\theta$$

$$= \frac{K}{\Gamma(n+1)} \left[\int_0^{\theta_1} \frac{\exp\left(-\frac{S}{\theta}\right)}{\theta^n}\, d\theta - \int_0^{\theta_1} \frac{\exp\left(-\frac{S+nx_{(1)}}{\theta}\right)}{\theta^n}\, d\theta \right]$$

$$= \frac{K}{\Gamma(n+1)} \left[\frac{1}{S^{n-1}} \int_{\frac{S}{\theta_1}}^{\infty} \exp(-u)\, u^{n-2}\, du \right.$$

$$- \frac{1}{\{S + nx_{(1)}\}^{n-1}} \int_{\frac{\{S+nx_{(1)}\}}{\theta_1}}^{\infty} \exp(-u) \cdot u^{n-2}\, du \Bigg]$$

$$= \frac{K}{S^{n-1}\,\Gamma(n+1)} \left[\int_{\frac{S}{\theta_1}}^{\infty} u^{n-2} \exp(-u)\, du - \left(\frac{S}{n\bar{x}}\right)^{n-1} \right.$$

$$\left. \int_{\frac{n\bar{x}}{\theta_1}}^{\infty} u^{n-2} \exp(-u)\, du \right]$$

Thus, θ_1 is the solution of the equation

$$\frac{\alpha S^{n-1}\,\Gamma(n+1)}{2K} = \int_{\frac{S}{\theta_1}}^{\infty} u^{n-2} \exp(-u)\, du$$

$$- \left(\frac{S}{n\bar{x}}\right)^{n-1} \int_{\frac{n\bar{x}}{\theta_1}}^{\infty} \exp(-u)\, u^{n-2}\, du \qquad (4.10)$$

and similarly θ_2 is the solution of the equation

$$\frac{\alpha S^{n-1}\,\Gamma(n+1)}{2K} = \int_{0}^{\frac{S}{\theta_2}} \exp(-u)\, u^{n-2}\, du$$

$$- \left(\frac{S}{n\bar{x}}\right)^{n-1} \int_{0}^{\frac{n\bar{x}}{\theta_2}} \exp(-u)\, u^{n-2}\, du \qquad (4.11)$$

where K is given by (4.7).

Given the data $\underset{\sim}{x}$, one may plot the posterior distribution $\Pi_2(\theta \mid \underset{\sim}{x})$ and if the posterior is unimodal, the $(1-\alpha)$-highest posterior density interval $[H_1, H_2]$ for θ may be constructed as defined by (4.1) and (4.2), H_1 and H_2 must simultaneously satisfy the equations

$$\int_{\frac{S}{H_2}}^{\frac{S}{H_1}} y^{n-2} \exp(-y)\, dy - \left\{\frac{S}{S+nx_{(1)}}\right\}^{n-1}.$$

$$\int_{S+\frac{nx_{(1)}}{H_2}}^{\frac{S+nx_{(1)}}{H_1}} y^{n-2} \exp(-y)\, dy = \frac{(1-\alpha)\, S^{n-1}\, \Gamma(n+1)}{K} \tag{4.12}$$

and,

$$\exp\left\{-S\left(\frac{1}{H_1}-\frac{1}{H_2}\right)\right\}\left[\frac{1-\exp\left(-\dfrac{nx_{(1)}}{H_1}\right)}{1-\exp\left(-\dfrac{nx_{(1)}}{H_2}\right)}\right] = \left(\frac{H_1}{H_2}\right)^n. \tag{4.13}$$

Example 4.2

Using Grubb's (1971) data $n = 19$, $x_{(1)} = 162$, and $S = 15{,}869$ we will compute 95% credible for μ and Credible/HPD intervals for θ.

From (4.8) and (4.9) at $\alpha = 0.05$ we obtain

$$\left[C_L^{(\mu)},\ C_U^{(\mu)}\right] \equiv [25.1187,\ 160.8734] \qquad \text{(Sinha, 1986)}$$

From (4.10) – (4.13) we obtain

$$\left[C_L^{(\theta)},\ C_U^{(\theta)}\right] = [581.05,\ 1466.33]$$

$$\left[H_L^{(\theta)},\ H_U^{(\theta)}\right] = [536.93,\ 1381.75] < \left[C_L^{(\theta)},\ C_U^{(\theta)}\right].$$

since $\quad H_U^{(\theta)} - H_L^{(\theta)} = 844.82 < C_U^{(\theta)} - C_L^{(\theta)} = 885.28$

4.3 NORMAL DISTRIBUTION

From (2.5) and (2.6) we have the marginal posteriors of μ and σ^2 with respect to the prior $g(\mu, \sigma) \propto \frac{1}{\sigma}$, viz.

$$\Pi_1(\mu/\underline{x}) = \frac{\sqrt{\frac{n}{A}}}{B\left(\frac{1}{2}, \frac{n-1}{2}\right)\left\{1+\frac{n(\bar{x}-\mu)^2}{A}\right\}^{n/2}}, \quad -\infty < \mu < \infty \tag{4.14}$$

and,

$$\Pi_2(\sigma^2 \backslash \underline{x}) = \frac{\left(\frac{A}{2}\right)^{\frac{n-1}{2}} \exp\left(-\frac{A}{2\sigma^2}\right)}{\Gamma\left(\frac{n-1}{2}\right)(\sigma^2)^{\frac{n+1}{2}}}, \quad \sigma > 0. \tag{4.15}$$

From (4.14) it follows that

$$\frac{\sqrt{n-1}\,(\mu-\bar{x})}{\sqrt{\frac{A}{n}}} \sim \text{Student's-}t \text{ with } (n-1) \text{ degrees of freedom.}$$

Also, $\Pi_1(\mu \mid \underline{x})$ is unimodal and symmetric about the mean $\bar{x}$. Hence, the $(1-\alpha)$-equal tail, the shortest credible and the HPD intervals for μ are identical. Such an interval (L, U) must satisfy the condition,

$$\begin{aligned} 1-\alpha &= P(L < \mu < U) \\ &= P\left[\frac{\sqrt{n-1}\,(L-\bar{x})}{\sqrt{\frac{A}{n}}} < t < \frac{\sqrt{n-1}\,(U-\bar{x})}{\sqrt{\frac{A}{n}}}\right] \\ &= P\left[\frac{L-\bar{x}}{s/\sqrt{n}} < t < \frac{U-\bar{x}}{s/\sqrt{n}}\right] \\ &\equiv P\left[-t\left(\frac{\alpha}{2}, n-1\right) < t < t\left(\frac{\alpha}{2}, n-1\right)\right]. \end{aligned}$$

Hence, we have

$$\left.\begin{aligned} L &= \bar{x} - \frac{s}{\sqrt{n}}\, t\left(\frac{\alpha}{2}, n-1\right) \\ U &= \bar{x} + \frac{s}{\sqrt{n}}\, t\left(\frac{\alpha}{2}, n-1\right) \end{aligned}\right\} \tag{4.16}$$

where $t(\alpha,\nu)$ = 100α % point of student's-t distribution with ν degrees of freedom.

Note that the $(1-\alpha)$-HPD interval of μ is the same as the classical confidence counterpart.

From (4.15) it follows that $\frac{A}{\sigma^2}$ is distributed as a χ^2 with $(n-1)$ degrees of freedom.

The $(1-\alpha)$-equal-tail credible interval $[T_1, T_2]$ must satisfy

$$\frac{\alpha}{2} = \int_0^{T_1} \Pi(\sigma^2 \mid \underline{x})\, d\sigma^2 = \int_{T_2}^{\infty} \Pi(\sigma^2 \mid \underline{x})\, d\sigma^2$$

$$\frac{\alpha}{2} = P(\sigma^2 \le T_1)$$

$$= P\left[\frac{A}{\sigma^2} \ge \frac{A}{T_1}\right]$$

$$= P\left[\chi^2 \ge \frac{A}{T_1} = \chi^2\left(\frac{\alpha}{2}, n-1\right)\right].$$

Similarly $$1-\frac{\alpha}{2} = P\left[\chi^2 \le \frac{A}{T_2} = \chi^2\left(1-\frac{\alpha}{2}, n-1\right)\right].$$

Thus, the $(1-\alpha)$-equal-tail credible interval of σ^2 is

$$[T_1, T_2] = \left[\frac{A}{\chi^2\left(\frac{\alpha}{2}, n-1\right)}, \frac{A}{\chi^2\left(1-\frac{\alpha}{2}, n-1\right)}\right] \tag{4.17}$$

where $\chi^2(\alpha, \nu)$ = upper 100α % point of a χ^2-distribution with ν degrees of freedom.

We observe that the credible interval (T_1, T_2) is the same as the corresponding classical confidence interval (Sinha and Howlader, 1984).

We leave the credible/HPD interval for lognormal parameters as an exercise for the readers.

4.4 WEIBULL DISTRIBUTION

Consider the Weibull density function

$$f(x \mid \theta, p) = \frac{p}{\theta} x^{p-1} \exp\left(-\frac{x^p}{\theta}\right), \quad p, \theta, x > 0.$$

From (2.40), the marginal posterior distribution of p is given by

$$\Pi_1 \,(p \mid \underline{x}) = K_1 \, p^{n-1} \, \lambda^{p-1} \Big/ \left(\sum x_i^p\right)^n \tag{4.18}$$

where

$$K_1^{-1} = \int_0^\infty p^{n-1} \, \lambda^{p-1} \Big/ \left(\sum x_i^p\right)^n dp.$$

Similarly, integrating out p from (2.40) the marginal posterior of θ is given by

$$\Pi_2 \,(\theta \mid \underline{x}) = K_2 \, \theta^{-(n+1)} \int_0^\infty p^{n-1} \, \lambda^{p-1} \exp\left(-\sum_i x_i^p / \theta\right) dp \tag{4.19}$$

where $K_2 = \Gamma(n)/K_1$.

The $(1-\alpha)$-credible interval for p is given by

$$\left[C_1^{(p)}, C_2^{(p)}\right] = [p_1, p_2],$$

where
$$\frac{\alpha}{2} = \int_0^{p_1} \Pi_1 \,(p \mid \underline{x}) \, dp = \int_{p_2}^\infty \Pi_1 \,(p \mid \underline{x}) \, dp. \tag{4.20}$$

If $\Pi_1 \,(p \mid \underline{x})$ is unimodal, the corresponding HPD interval $[H_1^{(p)}, H_2^{(p)}] \equiv [H_1, H_2]$ where H_1 and H_2 simultaneously satisfy

$$\Pi_1 \,(H_1 \mid \underline{x}) = \Pi_1 \,(H_2 \mid \underline{x}) \tag{4.21}$$

and,

$$\int_{H_1}^{H_2} \Pi_1 \,(p \mid \underline{x}) \, dp = 1 - \alpha. \tag{4.22}$$

Similarly, for the parameter θ, the $(1-\alpha)$-credible interval for θ is given by

$$\left[C_1^{(\theta)}, C_2^{(\theta)}\right] \equiv \left[\theta_1, \theta_2\right],$$

where $$\int_0^{\theta_1} \Pi_2\ (\theta \mid \underline{x})\ d\theta = \int_{\theta_2}^{\infty} \Pi_2\ (\theta \mid \underline{x})\ d\theta = \frac{\alpha}{2} \qquad (4.23)$$

and the corresponding HPD interval

$$\left[H_1^{(\theta)},\ H_2^{(\theta)}\right] \equiv [h_1, h_2]\ ,$$

where $$\Pi_2\ (h_1 \mid \underline{x}) = \Pi_2\ (h_2 + \underline{x})\ , \qquad (4.24)$$

and $$\int_{h_1}^{h_2} \Pi_2\ (\theta \mid \underline{x})\ d\theta = 1 - \alpha \qquad (4.25)$$

if, of course, $\Pi_2\ (\theta \mid \underline{x})$ is unimodal.

Example 4.3

Using Sinha (1986, p/53, Ex. 2.2) data:

$n = 25$, $\lambda = 4317.01$, $\alpha = 0.05$, from (4.20)–(4.25) we have 95% credible/HPD intervals

$$\left[C_1^{(p)},\ C_2^{(p)}\right] \equiv [1.457, 2.634]$$

$$\left[H_1^{(p)},\ H_2^{(p)}\right] \equiv [1.441, 2.611],\ H_2^{(p)} - H_1^{(p)} < C_2^{(p)} - C_1^{(p)}.$$

$$\left[C_1^{(\theta)},\ C_2^{(\theta)}\right] \equiv [1.899, 6.614]$$

$$\left[H_1^{(\theta)},\ H_2^{(\theta)}\right] \equiv [1.647, 6.085],\ H_2^{(\theta)} - H_1^{(\theta)} < C_2^{(\theta)} - C_1^{(\theta)}.$$

(Sinha and Guttman, 1988).

4.5 RAYLEIGH DISTRIBUTION

Given a random sample $\underline{x} = (x_1, x_2 \cdots, x_n)$ from the Rayleigh pdf,

$$f(x \mid \sigma) = \frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right) \quad x, \sigma > 0 \qquad (4.26)$$

under Jeffreys (1961) prior $g(\sigma) \propto \frac{1}{\sigma}$, from (2.47) we have obtained the posterior distribution

$$\Pi\ (\sigma \mid \underline{x}) = \frac{K}{\sigma^{2n+1}} \exp\left(-\frac{S^2}{2\sigma^2}\right) \qquad (4.27)$$

where $$S^2 = \sum_{i=1}^{n} x_i^2, \quad K = \frac{2\left(\frac{S^2}{2}\right)^n}{\Gamma(n)}.$$

From (4.27) it follows that

$$\frac{S^2}{\sigma^2} \sim \chi^2 (2n). \text{ Hence,}$$

$$1 - \alpha = P\left[\chi^2\left(1 - \frac{\alpha}{2}, 2n\right) < \frac{S^2}{\sigma^2} < \chi^2\left(\frac{\alpha}{2}, 2n\right)\right]$$

$$= P\left[\frac{S^2}{\chi^2\left(\frac{\alpha}{2}, 2n\right)} < \sigma^2 < \frac{S^2}{\chi^2\left(-\frac{\alpha}{2}, 2n\right)}\right]$$

Thus, $$\left[C_L^{(\sigma)}, C_U^{(\sigma)}\right] \equiv \left[\frac{S}{\sqrt{\chi^2\left(\frac{\alpha}{2}, 2n\right)}}, \frac{S}{\sqrt{\chi^2\left(1-\frac{\alpha}{2}, 2n\right)}}\right] \tag{4.28}$$

is the $(1-\alpha)$-credible interval for σ.

Note that the posterior distribution $\Pi(\sigma \mid \underline{x})$ is unimodal. Thus, the $(1-\alpha)$-shortest credible interval for σ is the same as the corresponding HPD interval. The $(1-\alpha)$-HPD interval $(H_U, H_L) \equiv (H_U^{(\sigma)}, H_L^{(\sigma)})$ must simultaneously satisfy

$$1 - \alpha = P\left[\frac{S^2}{H_U^2} < \chi^2 (2n) < \frac{S^2}{H_L^2}\right] \tag{4.29}$$

$$\exp\left[-\frac{S^2}{2}\left(\frac{1}{H_L^2} - \frac{1}{H_U^2}\right)\right] = \left(\frac{H_L}{H_U}\right)^{2n+1} \tag{4.30}$$

Example 4.4

Let $x = (x_1, x_2, \cdots, x_n)$, $n = 15$ is a random sample from the Rayleigh pdf (4.26).

Data x:

4.9225,	1.8637,	9.7437,	2.5716,	5.0859
9.7849,	5.5530,	2.8264,	3.0413,	9.8272
10.8185,	5.8119,	6.0044,	12.7716,	5.4174

We will construct 95% credible and HPD interval for σ.

$$S^2 = \sum_{i=1}^{15} x_i^2 = 757.1789\,.$$

$$\chi^2\,(0.025,30) = 46.9792, \quad \chi^2\,(0.975,30) = 16.7898\,.$$

From (4.28) we have

$$\left[C_L^{(\sigma)},\ C_U^{(\sigma)}\right] \equiv \left[\sqrt{\frac{757.1789}{46.9792}},\ \sqrt{\frac{757.1789}{16.7898}}\right]$$

$$\equiv [4.0146,\ 6.7155].$$

From (4.29) and (4.30) we obtain

$$\left[H_L^{(\sigma)},\ H_U^{(\sigma)}\right] \equiv [3.9025,\ 6.5415],\ H_U^{(\sigma)} - H_L^{(\sigma)} < C_U^{(\sigma)} - C_L^{(\sigma)}\,.$$

We may check the solutions $[H_L, H_U]$.

From (4.29)

$$P\left[\frac{S^2}{H_U^2} < \chi^2\,(30) < \frac{S^2}{H_L^2}\right]$$

$$= P\left[\frac{757.1789}{(6.5415)^2} < \chi^2\,(30) < \frac{757.1789}{(3.9025)^2}\right]$$

$$= P\left[17.6947 < \chi^2\,(30) < 49.7179\right]$$

$$= 0.9867 - 0.0367 = 0.95\,.$$

Left hand side of (4.30)

$$= \exp\left[-\frac{757.1789}{2}\{(3.9025)^{-2} - (6.5415)^{-2}\}\right]$$

$$= 1.11 \times 10^{-7}$$

$$\text{Right hand side of (4.30)} = \left(\frac{3.9025}{6.5415}\right)^{31} = 1.11 \times 10^{-7}.$$

Thus, $[H_L, H_U]$ simultaneously satisfies (4.29) and (4.30).

EXERCISES

1. Let $\underline{x} = (x_1, x_2, \cdots, x_n)$, $n = 15$ be a random sample from the exponential pdf (4.3).

   (i) Plot the posterior distribution $\Pi(\theta \mid \underline{x})$.

   (ii) Compute 90% credible/HPD intervals for θ under the prior $g(\theta) \propto \dfrac{1}{\theta}$, $\theta > 0$.

   Data $\{x_i\}$:

0.0600	2.0437	2.4818	9.3731	0.6282
1.2421	3.7502	0.4057	4.1299	6.0035
0.7405	3.3484	9.9887	7.5306	5.6484

2. Use the data in Exercise 1 above and construct the credible/HPD intervals with respect to the prior $g(\theta) \propto \dfrac{1}{\theta^2}$, $\theta > 0$. How do the intervals compare with the ones obtained under the prior $g(\theta) \propto \dfrac{1}{\theta}$, $\theta > 0$?

3. Suppose 20 items were subjected to a life testing experiment and the following failure times (x_i), $i = 1, 2, \cdots, 20$ (in units of 100 hours) were recorded.

   Data $\{x_i\}$:

5.0077	5.4352	5.7342	6.4556
5.1095	5.4464	5.8241	6.9005
5.1855	5.5131	5.9258	7.3815
5.3508	5.6047	6.3226	8.1383
5.4325	5.6762	6.3550	8.6077

Assuming that the life distribution follows the left truncated exponential

$$f(x \mid \mu, \theta) = \frac{1}{\theta} \exp\left(-\frac{(x-\mu)}{\theta}\right), \quad 0 \leq \mu \leq x_{(1)} \quad \theta > 0,$$

compute 90% credible interval for μ and credible HPD intervals for θ with respect to the prior $g(\mu,) \propto \frac{1}{\theta}$.

4. Obtain 90%-credible and HPD intervals for (μ,θ) in the two-parameter exponential pdf

$$f(x \mid \mu, \theta) = \frac{1}{\theta} \exp\left(-\frac{(x-\mu)}{\theta}\right), \quad \infty < \mu < \infty, \quad \theta > 0,$$

under the prior $g(\mu,\theta) \propto \frac{1}{\theta}$, given the sample $\underline{x} = (x_1, x_2, \cdots, x_n)$ from $f(x \mid \mu, \theta)$ above.

5. Let $X \sim N(\mu, \sigma^2)$ and data $\underline{x}$.
Given $\underline{x} = (x_1, x_2, \cdots, x_n)$, derive $(1-\alpha)$-HPD intervals for μ and σ^2 under the prior $g(\mu, \sigma) \propto \frac{1}{\sigma}$.

6. Let $Y = \log X \sim N(\mu, \sigma^2)$.
Given $\underline{x} = (x_1, x_2, \cdots, x_n)$, derive $(1-\alpha)$-HPD intervals for μ and σ^2 under the prior $g(\mu, \sigma) \propto \frac{1}{\sigma}$.

7. Given the following data $\underline{x}$ from the lognormal distribution in Exercise 6, compute 90% HPD intervals for μ and σ^2 under the same prior $g(\mu, \sigma) \propto \frac{1}{\sigma}$.

Data x:

16.38	40198.67	26.05	1549.81
0.28	630.81	5.02	43.68
356.43	19.77	1448.26	7.65
16.48	3.768	26.17	15.26
1.09	681.73	6.83	232.40

8. Let $(x_1, x_2, \cdots, x_n)$, $n = 15$ be a random sample from the Weibull pdf,

$$f(x \mid \theta, p) = \frac{p}{\theta} x^{p-1} \exp\left(-\frac{x^p}{\theta}\right), \quad x, p, \theta > 0 .$$

Compute 90% credible/HPD intervals for p and θ with respect to Jeffreys (1961) prior $g(\theta, p) \propto \dfrac{1}{p\theta}$.

Data: x

0.3749	2.8521	1.2159	2.7764	2.5382
0.5887	1.6222	2.3232	1.6370	2.4468
0.9070	0.9527	1.7148	1.5194	4.0344

9. Let $\underline{x} = (x_1, x_2, \cdots, x_n)$, $n = 15$ be a random sample from the Rayleigh pdf (4.26). Plot the posterior distribution $\Pi(\sigma \mid \underline{x})$ in (4.27) under Jeffreys (1961) prior $g(\sigma) \propto \dfrac{1}{\sigma}$ and compute 90% credible/HPD intervals for σ.

Data : x

2.9662	1.7335	4.1820	7.0800	5.7438
9.0979	6.0160	10.4881	4.6495	7.3655
6.5766	1.8139	6.9519	2.2676	7.5097

10. (a) Derive $\Pi(\sigma \mid \underline{x})$ with respect to ALI prior $g(\sigma) \propto \dfrac{1}{\sigma^3}$ (Hartigan, 1964) for the above pdf.

(b) Use the data above to compute 90% credible/HPD intervals for σ. How do the intervals compare with those obtained under Jeffreys (1961) prior?

CHAPTER

5 Bayesian Approximation of Posterior Moments and their Applications

5.1 INTRODUCTION

Bayes estimators are often obtained as a ratio of two integral expressions that cannot be expressed in a closed form. For some distributions, with appropriately chosen conjugate priors, the estimators may be routinely obtained in relatively simple forms, but in most cases numerical approximations are necessary — particularly if the parameters of interest are multidimensional Gauss-Hermite quadrature formula (Naylor and Smith, 1983) and Monte-Carlo integration routine (Kloek and Van Dijik, 1978; Zellner and Ross, 1984) are more accurate than the usual normal approximation but these approximations are computationally intensive and unattractive (Tierney and Kadane, 1986).

Lindley (1980) considered the ratio of two integrals.

$$I = \frac{\int_{\Omega} w(\theta) \ \exp\{L(\theta)\} \ d\theta}{\int_{\Omega} v(\theta) \ \exp\{L(\theta)\} \ d\theta}$$

where $\theta = (\theta_1, \theta_2, \cdots, \theta_m)$ is the parameter of interest, $L(\theta)$ is the logarithm of the likelihood function, $w(\theta)$ and $v(\theta)$ are arbitrary func-

tions of θ, and Ω is the parameter space of θ. Let $w(\theta) = u(\theta)\ v(\theta)$, and let $v(\theta)$ be the prior distribution of θ. We then have

$$I = \frac{\int_\Omega u(\theta)\ v(\theta)\ \exp\{L(\theta)\}\ d\theta}{\int_\Omega v(\theta)\ \exp\{L(\theta)\}\ d\theta}$$

$$= \frac{\int_\Omega u(\theta)\ v(\theta)\ l(\theta \mid \underline{x})\ d\theta}{\int_\Omega v(\theta)\ l(\theta \mid \underline{x})\ d\theta}$$

$$= E\{u(\theta) \mid \underline{x}\}$$

where $l(\theta \mid \underline{x})$ is the likelihood function and $\underline{x} = (x_1, x_2, \cdots, x_n)$ is a random sample from a distribution characterized by the density function $f(x \mid \theta)$.

In this chapter we discuss two methods of approximating the posterior moment,

$$E\{u(\theta) \mid \underline{x}\} = \frac{\int_\Omega u(\theta)\ v(\theta)\ \exp\{L(\theta)\}\ d\theta}{\int_\Omega v(\theta)\ \exp\{L(\theta)\}\ d\theta} \tag{5.1}$$

viz., Lindley's approximation (1980) and Tierney and Kadane's (T–K) approximation, (1986).

5.2 LINDLEY'S APPROXIMATION

The basic idea in Lindley's approach is to obtain Taylor series expansion of the functions involved in (5.1) about the maximum likelihood estimator $\hat{\theta}$ (see Sinha, 1986 for detailed derivation). Lindley approximated (5.1) by

$$E\{u(\theta) \mid \underline{x}\} = \left[\begin{array}{l} u + \frac{1}{2} \sum_i \sum_j (u_{ij} + 2u_i \rho_j)\ \sigma_{ij} \\ + \frac{1}{2} \sum_i \sum_j \sum_k \sum_r L_{ijk}\ \sigma_{ij}\ \sigma_{kr} \end{array}\right] + 0\left(\frac{1}{n^2}\right) \tag{5.2}$$

where, $i, j, k, r = 1, 2, \cdots, m$;

$\theta = (\theta_1, \theta_2, \cdots, \theta_m)$, $\hat{\theta}$ is the MLE of θ,

$$u_i = \frac{\partial u}{\partial \theta_i},\ u_{ij} = \frac{\partial^2 u}{\partial \theta_i\, \partial \theta_j},\ L_{ijk} = \frac{\partial^3 L}{\partial \theta_i\, \partial \theta_j\, \partial \theta_k}$$

$$\rho = \rho(\theta) = \log\ v(\theta),\ \rho_i = \frac{\partial \rho}{\partial \theta_i},$$

and σ_{ij} is the $(i, j)^{\text{th}}$ element in $[-L_{ij}]^{-1}$.

(5.2) is a "very good operational approximation of the ratio of multi-dimensional integrals" (Gren, 1980).

For a single-parameter case $m = 1$, (5.2) reduces to

$$E(u \mid \underline{x}) = u + \frac{1}{2}(u_{11} + 2u_1\rho_1)\sigma^2 + \frac{\sigma^4}{2} L_{111} u_1 + 0\left(\frac{1}{n^2}\right)$$

$$= \left[u + \frac{1}{2}(u_2 + 2u_1\rho_1)\sigma^2 + \frac{\sigma^4}{2} L_3 u_1 + 0\left(\frac{1}{n^2}\right)\right]_{\hat{\theta}}. \tag{5.3}$$

where we write $u_2 = u_{11} = \dfrac{\partial^2 u}{\partial \theta^2}$,

$L_3 = L_{111}$ which implies 'differentiate L three times with respect to θ

$$= \frac{\partial^3 L}{\partial \theta^3}$$

and
$$\sigma^2 = \left[-\frac{\partial^2 L}{\partial \theta^2}\right]^{-1}.$$

Note that under squared-error loss function, $E\{u(\theta) \mid \underline{x}\} = u^*$ is the Bayes estimator of $u(\theta)$. We now apply Lindley's approximation to the following distributions.

Binomial

We will obtain Bayes estimator of p under ALI prior $v(p) \propto \dfrac{1}{p(1-p)}$ (Hartigan, 1964) using the approximation (5.3) for $m = 1$.

We have $u = p,\ u_1 = 1,\ u_2 = 0,\quad \rho \propto -[\log p + \log(1-p)]$

$$L = Const + x \log p + (n-x) \log(1-p)$$

$$\frac{\partial^2 L}{\partial p^2} = -\frac{x}{p^2} - \frac{n-x}{(1-p)^2}$$

$$L^3 = \frac{\partial^3 L}{\partial p^3} = \frac{2x}{p^3} - \frac{2(n-x)}{(1-p)^3} \text{ and } \rho_1 = \frac{2p-1}{p(1-p)}.$$

At $p = \hat{p} = x/n,$

$$\frac{\partial^2 L}{\partial p^2} = -\frac{n}{p(1-p)},\ \sigma^2 = \frac{p(1-p)}{n}.$$

$$L_3 = 2n\left[\frac{1}{p^2} - \frac{1}{(1-p)^2}\right] = \frac{2n(1-2p)}{p^2(1-p)^2}.$$

Substituting in (5.3)

$$p^* = \left[p + \frac{2p-1}{p(1-p)} \cdot \frac{p(1-p)}{n} + \frac{n(1-2p)}{p^2(1-p)^2} \cdot \frac{p^2(1-p)^2}{n^2} + 0\left(\frac{1}{n^2}\right)\right]_{p=\hat{p}}$$

$$= \hat{p} + 0\left(\frac{1}{n^2}\right)$$

$$= \frac{x}{n} + 0\left(\frac{1}{n^2}\right). \tag{5.4}$$

Exponential

Consider the exponential pdf,

$$f(x \mid \theta) = \frac{1}{\theta} \exp\left(-\frac{x}{\theta}\right) \quad x, \theta > 0.$$

We derive Bayes estimator of θ with respect to ALI prior (Hartigan, 1964) $\nu(\theta) \propto \frac{1}{\theta^2}$ using the approximation (5.3).

We have $u = \theta,\ u_1 = 1,\ u_2 = 0$

$$\rho \propto -2 \log \theta,\ \rho_1 = -\frac{2}{\theta}$$

$$L = -n \log \theta - \frac{n\bar{x}}{\theta}$$

$$L_2 = \frac{n}{\theta^2} - \frac{2n\bar{x}}{\theta^3}$$

$$L_3 = -\frac{2n}{\theta^3} + \frac{6n\bar{x}}{\theta^4}.$$

At $\quad \theta = \hat{\theta} = \bar{x},\ L_2 = -\frac{n}{\theta^2},\ \sigma^2 = \frac{\theta^2}{n},$ and $L_3 = \frac{4n}{\theta^3}.$

Substituting in (5.3)

$$\theta^* = \left[\theta - \frac{2\theta^2}{n\theta} + \frac{2n\theta^4}{n^2\theta^2} + 0\left(\frac{1}{n^2}\right)\right]_{\hat{\theta}}$$

$$= \hat{\theta} + 0\left(\frac{1}{n^2}\right)$$

$$= \bar{x} + 0\left(\frac{1}{n^2}\right). \tag{5.5}$$

We now consider the two-parameter case, $m = 2,\ \theta = (\theta_1, \theta_2)$.

From (5.2) we have

$$\frac{1}{2}\sum_i \sum_j (u_{ij} + 2u_i\rho_j)\,\sigma_{ij}$$

$$= \frac{1}{2}(u_{11}\sigma_{11} + u_{12}\sigma_{12} + u_{21}\sigma_{21} + u_{22}\sigma_{22})$$

$$+ \rho_1(u_1\sigma_{11} + u_2\sigma_{21}) + \rho_2(u_1\sigma_{12} + u_2\sigma_{22}) \tag{5.6}$$

$$\sum_i \sum_j \sum_k \sum_r L_{ijk}\,\sigma_{ij}\,\sigma_{kr}\,u_r$$

$$= \sum_i \sum_j \sum_k L_{ijk}\,\sigma_{ij}\,(\sigma_{k1}u_1 + \sigma_{k2}u_2)$$

$$= \sum_i \sum_j \sigma_{ij}\left[L_{ij1}(\sigma_{11}u_1 + \sigma_{12}u_2) + L_{ij2}(\sigma_{21}u_1 + \sigma_{22}u_2)\right]$$

$$= (u_1\sigma_{11} + u_2\sigma_{12})(\sigma_{11}L_{111} + \sigma_{12}L_{121} + \sigma_{21}L_{211} + \sigma_{22}L_{221})$$

$$+ (u_1\sigma_{21} + u_2\sigma_{22})(\sigma_{11}L_{112} + \sigma_{12}L_{122} + \sigma_{21}L_{212} + \sigma_{22}L_{222}). \tag{5.7}$$

where, L_{112} implies 'differentiate L twice with respect to θ_1 and once with respect to θ_2' which we write $L_{112} = \frac{\partial^3 L}{\partial\theta_1^2\,\partial\theta_2}$ etc.

Clearly $\quad L_{111} = L_{30} = \dfrac{\partial^3 L}{\partial \theta_1^3}$,

$$L_{112} = L_{121} = L_{211} = L_{21} = \frac{\partial^3 L}{\partial \theta_1^2\, \partial \theta_2},$$

$$L_{122} = L_{212} = L_{221} = L_{12} = \frac{\partial^3 L}{\partial \theta_1\, \partial \theta_2^2}.$$

Substituting in (5.7)

$$\sum_i \sum_j \sum_k \sum_r L_{ijk}\, \sigma_{ij}\, \sigma_{kr}\, u_r$$

$$= (u_1\sigma_{11} + u_2\sigma_{12})\,(\sigma_{11}L_{30} + 2\sigma_{12}L_{21} + \sigma_{22}L_{12})$$

$$+ (u_1\sigma_{21} + u_2\sigma_{22})\,(\sigma_{11}L_{21} + 2\sigma_{12}L_{12} + \sigma_{22}L_{03})$$

$$= (u_1\sigma_{11}^2 + u_2\sigma_{12}\sigma_{11})\, L_{30} + (2u_1\, \sigma_{11}\sigma_{12} + 2u_2\sigma_{12}^2$$

$$+ u_1\sigma_{21}\sigma_{11} + u_2\, \sigma_{22}\sigma_{11})\, L_{21} + (u_1\sigma_{11}\sigma_{22}$$

$$+ u_2\sigma_{12}\sigma_{22} + 2u_1\sigma_{12}^2 + 2u_2\sigma_{22}\, \sigma_{12})\, L_{12}$$

$$+ (u_1\sigma_{21}\sigma_{22} + u_2\sigma_{22}^2)\, L_{03}$$

$$= 3\,(u_2\sigma_{12}\sigma_{22}L_{12} + u_1\sigma_{11}\sigma_{12}L_{21}) + (\sigma_{11}\sigma_{22} + 2\sigma_{12}^2)\,(u_1L_{12} + u_2L_{21})$$

$$+ u_1\,(\sigma_{11}^2 L_{30} + \sigma_{21}\sigma_{22}L_{03}) + u_2\,(\sigma_{11}\sigma_{12}L_{30} + \sigma_{22}^2 L_{03}). \tag{5.8}$$

Substituting (5.6) and (5.8) in (5.2),

$$u^* = \Bigg[u + \frac{1}{2}\,(u_{11}\sigma_{11} + u_{12}\sigma_{12} + u_{21}\sigma_{21} + u_{22}\sigma_{22})$$

$$+ \rho_1\,(u_1\sigma_{11} + u_2\sigma_{21}) + \rho_2\,(u_1\sigma_{21} + u_2\sigma_{22})$$

$$+ \frac{1}{2}\Big\{3\,(u_2\sigma_{12}\sigma_{22}L_{12} + u_1\sigma_{11}\sigma_{12}L_{21})$$

$$+ (\sigma_{11}\sigma_{22} + 2\sigma_{12}^2)\,(u_1L_{12} + u_2L_{21})$$

$$+ u_1\,(\sigma_{11}^2 L_{30} + \sigma_{21}\sigma_{22}L_{03}) + u_2\,(\sigma_{11}\sigma_{12}L_{30} + \sigma_{22}^2 L_{03})$$

$$+ 0\left(\frac{1}{n^2}\right)\Bigg]_{\theta}. \tag{5.9}$$

If θ_1 and θ_2 are orthogonal, $\sigma_{ij} = 0,\ i \neq j$ and (5.9) simplifies to

$$u^* = \Big[u + \frac{1}{2}(u_{11}\sigma_{11} + u_{22}\sigma_{22}) + \rho_1 u_1 \sigma_{11} + \rho_2 u_2 \sigma_{22}$$

$$+ \frac{1}{2}\left\{\sigma_{11}\sigma_{22}\,(u_1 L_{12} + u_2 L_{21}) + u_1\sigma_{11}^2 L_{30} + u_2\sigma_{22}^2 L_{03}\right\} + 0\left(\frac{1}{n^2}\right)\Big]_{\hat{\theta}}. \tag{5.10}$$

Normal

Let $X \sim N(\mu, \sigma^2)$, $v(\mu, \sigma) \propto \dfrac{1}{\sigma}$ and suppose μ and σ are orthogonal. We will derive (μ^*, σ^*) using the approximation (5.10).

Set $(\theta_1, \theta_2) \equiv (\mu, \sigma)$;

$$L = \text{Const}-n\log\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$L_{30} = \frac{\partial^3 L}{\partial\mu^3} = 0, \quad L_{12} = \frac{\partial^3 L}{\partial\mu\,\partial\sigma^2} = \frac{6\sum_i (x_i-\mu)}{\sigma^4}$$

$$L_{21} = \frac{\partial^3 L}{\partial\mu^2\,\partial\sigma} = \frac{2n}{\sigma^3}$$

$$L_{03} = \frac{\partial^3 L}{\partial\sigma^3} = \frac{-2n}{\sigma^3} + \frac{12}{\sigma^5}\sum_i (x_i-\mu)^2$$

$$\frac{\partial^2 L}{\partial\mu^2} = \frac{-n}{\sigma^2}, \quad \frac{\partial^2 L}{\partial\mu\,\partial\sigma} = \frac{-2}{\sigma^3}\sum_i (x_i-\mu)$$

$$\frac{\partial^2 L}{\partial\sigma^2} = \frac{n}{\sigma^2} - \frac{3}{\sigma^4}\sum_i (x_i-\mu)^2.$$

At the MLE $(\hat{\mu}, \hat{\sigma})$ we have

$$-L_{ij} = \begin{bmatrix} \frac{n}{\sigma^2} & 0 \\ & \frac{2n}{\sigma^2} \\ 0 & \sigma^2 \end{bmatrix}, \quad \sigma_{11} = \frac{\sigma^2}{n},$$

$$\sigma_{22} = \frac{\sigma^2}{2n} \text{ and } L_{30} = 0,\ L_{03} = \frac{10n}{\sigma^3},\ L_{21} = \frac{2n}{\sigma^3},\ L_{12} = 0.$$

Substituting in (5.9), we have

$$\mu^* = \bar{x} + 0\left(\frac{1}{n^2}\right) \tag{5.11}$$

$$\sigma^* = \left[\sigma + \frac{\sigma^2}{2n}\left(-\frac{1}{\sigma}\right) + \frac{1}{2}\left(\frac{\sigma^4}{2n^2}\cdot\frac{2n}{\sigma^3} + \frac{\sigma^4}{4n^2}\cdot\frac{10n}{\sigma^3}\right) + 0\left(\frac{1}{n^2}\right)\right]_{\hat{\sigma}}$$

$$= \hat{\sigma}\left(1 + \frac{5}{4n}\right) + 0\left(\frac{1}{n^2}\right) \tag{5.12}$$

where $\hat{\sigma}^2 = \frac{\sum_i (x_i - \bar{x})^2}{n}$

Weibull

Consider Weibull pdf,

$$f(x \mid p, \theta) = \frac{p}{\theta} x^{p-1} \exp\left(-\frac{x^p}{\theta}\right), \quad p,\ \theta,\ x > 0. \tag{5.13}$$

Reliability function at any time t is given by

$$R_t = P(X > t) = \int_t^{\infty} f(x \mid \theta, p)\, dx$$

$$= \exp\left(-\frac{t^p}{\theta}\right).$$

We will obtain Bayes estimator of R_t with respect to Jeffreys (1961) prior $v(p, \theta) \propto \frac{1}{p\theta}$ using the approximation (5.9).

Assume p and θ are orthogonal

$$\sigma_{ij} = 0, \; i \neq j = 1, 2.$$

$$u \equiv \exp\left(-\frac{t^p}{\theta}\right).$$

$$u_1 = \frac{\partial u}{\partial \theta} = \frac{ut^p}{\theta^2}, \; u_{11} = \frac{\partial^2 u}{\partial \theta^2} = \frac{ut^p}{\theta^3}\left(\frac{t^p}{\theta} - 2\right)$$

$$\rho \propto -\log p - \log \theta, \; \rho_1 = -\frac{1}{\theta}, \; \rho_2 = -\frac{1}{p}.$$

$$L = n \log p - n \log \theta - \frac{\sum_i x_i^p}{\theta} + (p-1)\sum_i \log x_i.$$

$$\frac{\partial L}{\partial \theta} = -\frac{n}{\theta} + \frac{\sum_i x_i^p}{\theta^2}$$

$$\frac{\partial L}{\partial p} = \frac{n}{p} - \frac{\sum_i x_i^p \log x_i}{\theta} + \sum_i \log x_i.$$

$$L_{20} = \frac{\partial^2 L}{\partial \theta^2} = \frac{n}{\theta^2} - \frac{2\sum_i x_i^p}{\theta^3}$$

$$L_{02} = \frac{\partial^2 L}{\partial p^2} = -\frac{n}{p^2} - \frac{\sum_i x_i^p (\log x_i)^2}{\theta}$$

$$L_{21} = \frac{\partial^3 L}{\partial \theta^2 \partial p} = \frac{-2\sum_i x_i^p \log x_i}{\theta^3}$$

$$L_{12} = \frac{\partial^3 L}{\partial \theta \partial p^2} = \frac{\sum_i x_i^p (\log x_i)^2}{\theta^2}$$

all functions are to be evaluated at the MLE $(\hat{\theta}, \hat{p})$ where $(\hat{\theta}, \hat{p})$ are the solutions of the equations $\frac{\partial L}{\partial \theta} = 0$ and $\frac{\partial L}{\partial p} = 0.$

Cohen (1965) and Sinha and Kale (1980), have discussed the solutions of the above equations in details.

Note that at $(\hat{\theta}, \hat{p})$

$$\sigma_{22} = (L_{02})^{-1}, \ \sigma_{11} = (L_{20})^{-1} = \frac{n}{\theta^2}.$$

Given the data $\underline{x} = (x_1, x_2, \cdots, x_n)$ we obtain $(\hat{\theta}, \hat{p})$, and substituting in (5.10), Bayes estimator $R_{t_0}^*$ at a specified $t = t_0$ may be computed.

Inverse Gaussian

Consider the inverse Gaussian IG (λ, μ). Padgett (1981), Folks and Chhikara (1978), Whitmore (1976), Banerjee and Bhattacharyya (1977), Lancaster (1972) and Tweedie (1957), among others, have done extensive work on the theory and application of the Inverse Gaussian distribution in a wide variety of fields. Chhikara and Folks (1977) proposed the Inverse Gaussian as a life-time model and suggested its application for studying reliability aspects where the initial failure rate is high, and Nadas (1973) considered the Inverse Gaussian as a model in the context of first passage time in Brownian motion.

The pdf of the IG (λ, μ) random variable X is given by

$$f(x \mid \lambda, \mu) = \left(\frac{\lambda}{2\pi x^3}\right)^{1/2} \cdot \exp\left\{-\frac{\lambda\,(x-\mu)^2}{2\mu^2 x}\right\}, \quad x \geq 0,\ \lambda > 0,\ \mu > 0. \tag{5.14}$$

The MLE of μ and λ are

$$\hat{\mu} = \bar{x},\ \hat{\lambda} = \frac{n}{\sum_{i=1}^{n}\left(\frac{1}{x_i} - \frac{1}{\bar{x}}\right)} \qquad \text{(Tweedie, 1957)}$$

We will obtain Bayes estimators (μ^*, λ^*) using the approximation (5.10) and the prior $\nu(\lambda, \mu) \propto \frac{1}{\lambda}$.

$$L = \text{Constant} + \frac{n}{2}\log\lambda - \frac{\lambda}{2\mu^2}\left(n\bar{x} - 2n\mu + \mu^2 \sum_{i=1}^{n} \frac{1}{x_i}\right)$$

At the MLE $(\hat{\lambda}, \hat{\mu})$

$$L_{12} = \frac{\partial^3 L}{\partial\mu\,\partial\lambda^2} = 0, \quad L_{21} = \frac{\partial^3 L}{\partial\mu^3\,\partial\lambda} = -\frac{n}{\mu^3}$$

$$L_{30} = \frac{\partial^3 L}{\partial\mu^3} = \frac{6n\lambda}{\mu^4}, \quad L_{03} = \frac{\partial^3 L}{\partial\lambda^3} = \frac{n}{\lambda^3}$$

$$\rho = \log v(\lambda, \mu) \propto -\log\lambda$$

$$\rho_1 = \frac{\partial\rho}{\partial\mu} = 0, \quad \rho_2 = \frac{\partial\rho}{\partial\lambda} = -\frac{1}{\lambda}$$

$$\sigma_{12} = 0, \quad \sigma_{11} = \frac{\mu^3}{n\lambda}, \quad \sigma_{22} = \frac{2\lambda^2}{n}$$

Substituting in (5.10)

$$\mu^* = \left[\mu + \frac{1}{2} L_{30}\,\sigma_{11}^2 + 0\left(\frac{1}{n^2}\right)\right]_{(\hat{\mu}, \hat{\lambda})}$$

$$= \hat{\mu} + \frac{3\hat{\mu}^2}{n\hat{\lambda}} + 0\left(\frac{1}{n^2}\right)$$

$$\lambda^* = \left[\lambda - \frac{\sigma_{22}}{\lambda} + \frac{1}{2}\left(L_{03}\,\sigma_{22}^2 + L_{21}\sigma_{11}\sigma_{22}\right) + 0\left(\frac{1}{n^2}\right)\right]_{(\hat{\mu}, \hat{\lambda})}$$

$$= \left(\frac{n-1}{n}\right)\hat{\lambda} + 0\left(\frac{1}{n^2}\right).$$

Three-parameter Weibull

Consider a three-parameter Weibull pdf,

$$f(x \mid c, b, \mu) = \frac{c}{b}\left(\frac{x-\mu}{b}\right)^{c-1} \exp\left[-\left(\frac{x-\mu}{b}\right)^{c}\right],$$

$$b,\ c > 0,\ x > \mu \tag{5.15}$$

Suppose we are ignorant about the parameters (c, b, μ). Following Jeffreys' (1961), the vague priors,

$$g(b, c) \propto \frac{1}{bc} \quad \text{and} \quad h(\mu) = \text{constant}$$

would be appropriate.

It is reasonable to believe that μ is *a-priori* independent of b and c since any idea one may have about μ is not likely to influence very much one's knowledge about numerical values of b and c. Thus, we may write the joint prior distribution of (c, b, μ) as

$$v\ (c, b, \mu) \propto \frac{1}{bc}. \tag{5.16}$$

For $m = 3$, (5.2) reduces to

$$\begin{aligned} u^* &= E(u \mid \underline{x}) \\ &= u + (u_1 a_1 + u_2 a_2 + u_3 a_3 + a_4 + a_5) \\ &\quad + \frac{1}{2}\Big[A\,(u_1\sigma_{11} + u_2\sigma_{12} + u_3\sigma_{13}) + \\ &\quad + B\,(u_1\sigma_{21} + u_2\sigma_{22} + u_3\sigma_{23}) \\ &\quad + D\,(u_1\sigma_{31} + u_2\sigma_{32} + u_3\sigma_{33})\Big] + 0\left(\frac{1}{n^2}\right) \end{aligned} \tag{5.17}$$

evaluated at the MLE $\hat{\theta} = (\hat{c}, \hat{b}, \hat{\mu})$, where

$$a_1 = \rho_1\sigma_{11} + \rho_2\sigma_{12} + \rho_3\sigma_{13}$$

$$a_2 = \rho_1\sigma_{21} + \rho_2\sigma_{22} + \rho_3\sigma_{23}$$

$$a_3 = \rho_1\sigma_{31} + \rho_2\sigma_{32} + \rho_3\sigma_{33}, \ \rho_j = \frac{\partial v(\theta)}{\partial \theta_j};$$

$$a_4 = u_{12}\sigma_{12} + u_{13}\sigma_{13} + u_{23}\sigma_{23}$$

$$a_5 = \frac{1}{2}(u_{11}\sigma_{11} + u_{22}\sigma_{22} + u_{33}\sigma_{33}).$$

$$A = \sigma_{11}L_{111} + 2\sigma_{12}L_{121} + 2\sigma_{13}L_{131} + 2\sigma_{23}L_{231} + \sigma_{22}L_{221} + \sigma_{33}L_{331}$$

$$B = \sigma_{11}L_{112} + 2\sigma_{12}L_{122} + 2\sigma_{13}L_{132} + 2\sigma_{23}L_{232} + \sigma_{22}L_{222} + \sigma_{33}L_{332}$$

$$D = \sigma_{11}L_{113} + 2\sigma_{12}L_{123} + 2\sigma_{13}L_{133} + 2\sigma_{23}L_{233} + \sigma_{22}L_{223} + \sigma_{33}L_{333}$$

(Subscripts 1, 2, 3 refer to parameters c, b, μ, respectively).

For the prior distribution in (5.16), we have

$$\rho \propto -\log b - \log c.$$

$$\rho_1 = -\frac{1}{c},\ \rho_2 = -\frac{1}{b},\ \rho_3 = 0.$$

Define $u = c$.

$$u_1 = 1,\ u_2 = u_3 = 0,\ u_{ij} = 0,\ i, j = 1, 2, 3.$$

From (5.17) we obtain

$$c^* = \left[c - \left(\frac{\sigma_{11}}{c} + \frac{\sigma_{12}}{b}\right) + \frac{1}{2}\left(A\sigma_{11} + B\sigma_{12} + D\sigma_{13}\right)\right]_{\hat{\theta}}. \tag{5.18}$$

Define $u = b$.

$u_1 = 0 = u_3,\ u_2 = 1,$ and we have

$$b^* = \left[b - \left(\frac{\sigma_{21}}{c} + \frac{\sigma_{22}}{b}\right) + \frac{1}{2}\left(A\sigma_{12} + B\sigma_{22} + D\sigma_{23}\right)\right]_{\hat{\theta}}. \tag{5.19}$$

and, similarly

$$\mu^* = \left[\mu - \left(\frac{\sigma_{31}}{c} + \frac{\sigma_{32}}{b}\right) + \frac{1}{2}\left(A\sigma_{31} + B\sigma_{32} + D\sigma_{33}\right)\right]_{\hat{\theta}}. \tag{5.20}$$

The maximum likelihood estimators of the parameters of the Weibull density function (5.15) have been studied by several authors including Rockette, Antle and Klimko (1974) and Cohen and Whitten (1982). The 3-parameter Weibull distribution does not satisfy that usual regularity conditions and as a result for some combinations of parameter values and sample size the MLE does not exist or may lead to inconsistent estimators. Sinha and Sloan (1988) developed a modified Newton-Raphson iterative algorithm for computing the MLE $(\hat{c}, \hat{b}, \hat{\mu})$.

Note that $\sigma_{ij} = (i, j)^{\text{th}}$ element in the inverse of the negative Hessian, $i, j = 1, 2, 3$. Thus,

$$\begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}_\theta = - \begin{bmatrix} \dfrac{\partial^2 L}{\partial c^2} & \dfrac{\partial^2 L}{\partial c\, \partial b} & \dfrac{\partial^2 L}{\partial c\, \partial \mu} \\ \dfrac{\partial^2 L}{\partial c\, \partial b} & \dfrac{\partial^2 L}{\partial b^2} & \dfrac{\partial^2 L}{\partial b\, \partial \mu} \\ \dfrac{\partial^2 L}{\partial c\, \partial \mu} & \dfrac{\partial^2 L}{\partial \mu\, \partial b} & \dfrac{\partial^2 L}{\partial \mu^2} \end{bmatrix}_\theta \tag{5.21}$$

where,

$$L = n\,(\log c - c \log b) + (c - 1) \sum_{i=1}^{n} \log (x_i - \mu) - \sum_{i=1}^{n} \left(\frac{x_i - \mu}{b}\right)^c .$$

Reliability function at any time t is given by

$$R_t = P\,(X \geq t) = \int_t^\infty f(x \mid c, b, \mu)\; dx = \exp \left[- \left(\frac{t-\mu}{b}\right)^c \right], \quad t > \mu.$$

Putting $u = R_t = R$ in (5.11), we obtain

$$R^* = [R + \lambda_1\, U_1 + \lambda_2\, U_2 + \lambda_3\, U_3 + a_4 + a_5]_\theta \tag{5.22}$$

where,

$$\lambda_1 = \left(\frac{A}{2} - \frac{1}{c}\right), \; \lambda_2 = \left(\frac{B}{2} - \frac{1}{b}\right), \; \lambda_3 = \frac{D}{2}$$

$$U_1 = u_1\, \sigma_{11} + u_2\, \sigma_{12} + u_3\, \sigma_{13}$$

$$U_2 = u_1\, \sigma_{21} + u_2\, \sigma_{22} + u_3\, \sigma_{23}$$

$$U_3 = u_1\, \sigma_{31} + u_2\, \sigma_{32} + u_3\, \sigma_{33}\, .$$

Note that in this case u_i and $u_{ij} \neq 0, i, j = 1,2,3$.

Setting $u = R^2$, we have

$$E(R^2 \mid \underline{x}) = \begin{bmatrix} R^2 + 2R\,(\lambda_1\, U_1 + \lambda_2\, U_2 + \lambda_3\, U_3 + a_4 + a_5) \\ + u_1^2\, \sigma_{11} + u_2^2\, \sigma_{22} + u_3^2\, \sigma_{33} + 2u_1\, u_2\, \sigma_{12} \\ + 2u_1\, u_3\, \sigma_{13} + 2u_2\, u_3\, \sigma_{23} \end{bmatrix}_\theta \tag{5.23}$$

Using (5.22) and (5.23) we may compute the posterior variance

$$V(R \mid \underline{x}) = E(R^2 \mid \underline{x}) - \{E(R \mid \underline{x})\}^2, \quad \text{and}$$

similarly for $V(c\,|\,\underline{x})$, $V(b\,|\,\underline{x})$ and $V(\mu\,|\,\underline{x})$.

To evaluate the constants in the estimating equations we set up the $[\sigma_{ij}] = [-L_{ij}]^{-1}$ matrix given by (5.21) and the matrix,

$$[L_{ijk}]_{\hat\theta} = \begin{bmatrix} L_{111} & L_{112} & L_{113} \\ L_{221} & L_{222} & L_{223} \\ L_{331} & L_{332} & L_{333} \end{bmatrix}_{\hat\theta} = -\begin{bmatrix} \dfrac{\partial^3 L}{\partial c^3} & \dfrac{\partial^3 L}{\partial c^2\,\partial b} & \dfrac{\partial^3 L}{\partial c^2\,\partial\mu} \\ \dfrac{\partial^3 L}{\partial b^2\,\partial c} & \dfrac{\partial^3 L}{\partial b^3} & \dfrac{\partial^3 L}{\partial b^2\,\partial\mu} \\ \dfrac{\partial^3 L}{\partial\mu^2\,\partial c} & \dfrac{\partial^3 L}{\partial\mu^2\,\partial b} & \dfrac{\partial^3 L}{\partial\mu^3} \end{bmatrix}_{\hat\theta} \tag{5.24}$$

where,

$$\frac{\partial^2 L}{\partial c^2} = -\frac{n}{c^2} - \sum_{i=1}^{n}\left(\frac{x_i-\mu}{b}\right)^c\left[\log\left(\frac{x_i-\mu}{b}\right)\right]^2$$

$$\frac{\partial^2 L}{\partial b^2} = \frac{nc}{b^2} - \frac{c\,(c+1)}{b^{c+2}}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\frac{\partial^2 L}{\partial\mu^2} = -(c-1)\left[\sum_{i=1}^{n}(x_i-\mu)^{-2} + \frac{c}{b^2}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{b}\right)^{c-2}\right]$$

$$\frac{\partial^2 L}{\partial c\,\partial b} = -\frac{n}{b} + \frac{1}{b}\left[\sum_{i=1}^{n}\left(\frac{x_i-\mu}{b}\right)^c\left\{c\log\left(\frac{x_i-\mu}{b}\right)+1\right\}\right]$$

$$\frac{\partial^2 L}{\partial c\,\partial\mu} = -\sum_{i=1}^{n}(x_i-\mu)^{-1} + \frac{1}{b}\sum\left(\frac{x_i-\mu}{b}\right)^{c-1}\left[c\log\left(\frac{x_i-\mu}{b}\right)+1\right]$$

$$\frac{\partial^2 L}{\partial b\,\partial\mu} = \frac{-c}{b^2} - \sum_{i=1}^{n}\left(\frac{x_i-\mu}{b}\right)^{c-1}$$

$$L_{111} = \frac{\partial^3 L}{\partial c^3} = \frac{2n}{c^3} - \sum_{i=1}^{n}\left(\frac{x_i-\mu}{b}\right)^c\left[\log\left(\frac{x_i-\mu}{b}\right)\right]^3$$

$$L_{222} = \frac{\partial^3 L}{\partial b^3} = \frac{c}{b^3}\left[(c+1)\ (c+2)\ \sum_{i=1}^{n}\left(\frac{x_i-\mu}{b}\right)^c - 2n\right]$$

$$L_{223} = \frac{2c\,(c+1)}{b^{c+2}}\ \sum_{i=1}^{n}\ (x_i-\mu)$$

$$L_{333} = \frac{\partial^3 L}{\partial \mu^3} = (c-1)\left[\frac{c\,(c-2)}{b^3}\ \sum_{i=1}^{n}\left(\frac{x_i-\mu}{b}\right)^{c-3}\right.$$

$$\left. -2\ \sum_{i=1}^{n}\ (x_i-\mu)^{-3}\right]$$

$$L_{112} = \frac{\partial^3 L}{\partial c^2\,\partial b} = \frac{1}{b}\ \sum_{i=1}^{n}\left(\frac{x_i-\mu}{b}\right)^c$$

$$\left[c\left\{\log\left(\frac{x_i-\mu}{b}\right)\right\}^2 + 2\log\left(\frac{x_i-\mu}{b}\right)\right]$$

$$L_{221} = \frac{\partial^3 L}{\partial b^2\,\partial c} = \frac{n}{b^2} - \frac{1}{b^2}\ .$$

$$\left[\sum_{i=1}^{n}\left(\frac{x_i-\mu}{b}\right)^c \left\{(c^2+c)\log\left(\frac{x_i-\mu}{b}\right)+2c+1\right\}\right]$$

$$L_{132} = \frac{\partial^3 L}{\partial c\,\partial \mu\,\partial b} = -\frac{1}{b^2}\ .$$

$$\left[\sum_{i=1}^{n}\left(\frac{x_i-\mu}{b}\right)^{c-1} \left\{1+c\log\left(\frac{x_i-\mu}{b}\right)^c\right\}\right]$$

$$L_{113} = \frac{\partial^3 L}{\partial c^2\,\partial \mu} = \frac{1}{b}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{b}\right)^{c-1}\ .$$

$$\left[2\log\left(\frac{x_i-\mu}{b}\right)+c\left\{\log\left(\frac{x_i-\mu}{b}\right)\right\}^2\right]$$

$$L_{331}=\frac{\partial^3 L}{\partial\mu^2\,\partial c}=-\sum_{i=1}^{n}(x_i-\mu)^{-2}-\frac{1}{b^2}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{b}\right)^{c-2}\cdot$$

$$\left[c+(c-1)\left\{c\log\left(\frac{x_i-\mu}{b}\right)+1\right\}\right]$$

$$L_{332}=\frac{\partial^3 L}{\partial\mu^2\,\partial b}=\frac{c^2(c-1)}{b^3}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{b}\right)^{c-2}.$$

Reliability Estimation

Let $\quad u = R(t) = \exp\left\{-\left(\frac{t-\mu}{b}\right)^c\right\}, \quad t>\mu.$

$$u_1=\frac{\partial u}{\partial c}=-\left(\frac{t-\mu}{b}\right)^c\left\{\log\left(\frac{t-\mu}{b}\right)\right\}u$$

$$u_2=\frac{\partial u}{\partial b}=\left\{\frac{c}{b}\left(\frac{t-\mu}{b}\right)^c\right\}u$$

$$u_3=\frac{\partial u}{\partial \mu}=\left\{\frac{c}{b}\left(\frac{t-\mu}{b}\right)^{c-1}\right\}u$$

$$u_{11}=\frac{\partial^2 u}{\partial c^2}=u\left(\frac{t-\mu}{b}\right)^c\left[\left\{\log\left(\frac{t-\mu}{b}\right)\right\}^2\right]\left[\left(\frac{t-\mu}{b}\right)^c-1\right]$$

$$u_{12}=\frac{\partial^2 u}{\partial c\,\partial b}=\frac{u}{b}\left(\frac{t-\mu}{b}\right)^c.$$

$$\left[\left\{c\log\left(\frac{t-\mu}{b}\right)\right\}\left\{1-\left(\frac{t-\mu}{b}\right)^c\right\}+1\right]$$

$$u_{13}=\frac{\partial^2 u}{\partial c\,\partial\mu}=\frac{b}{t-\mu}\,u_{12}.$$

$$u_{22} = \frac{uc}{b^2}\left(\frac{t-\mu}{b}\right)^c\left[c\left(\frac{t-\mu}{b}\right)^c - c - 1\right] = \frac{\partial^2 u}{\partial b^2}$$

$$u_{23} = \frac{uc^2}{b^2}\left(\frac{t-\mu}{b}\right)^{c-1}\left[\left(\frac{t-\mu}{b}\right)^c - 1\right] = \frac{\partial^2 u}{\partial b\,\partial\mu}$$

$$u_{33} = \frac{\partial^2 u}{\partial\mu^2} = \frac{cu}{b^2}\left(\frac{t-\mu}{b}\right)^{c-2}\left[c\left(\frac{t-\mu}{b}\right)^c - c + 1\right].$$

Substituting in (5.18) – (5.24) we obtain $\theta^* = c^*, b^*, \mu^*)$ and $R*$. (Sinha and Sloan, 1988)

Example 5.1:

A random sample of size $n = 400$ was generated from the Weibull pdf (5.15) with $c = 3$, $b = 100$, $\mu = 30$.

$$\text{MLE } \hat{\theta} = (\hat{c}, \hat{b}, \hat{\mu}) \equiv (2.78330,\ 93.42635,\ 37.51172).$$

From (5.2), and (5.24)

$$[\sigma_{ij}]_{\hat\theta} = \begin{bmatrix} 0.06611 & 1.61250 & -1.42615 \\ 1.61250 & 47.74003 & -41.00272 \\ -1.42615 & -41.00272 & -37.67748 \end{bmatrix}$$

$$[L_{ijk}]_{\hat\theta} = \begin{bmatrix} 30.71913 & 2.25484 & 1.61270 \\ -0.32682 & 0.02180 & 0.01293 \\ -0.30190 & 0.00610 & -0.01284 \end{bmatrix}$$

$$A = -1.27420,\ B = 0.03480,\ D = -0.19940.$$

Substituting in (5.18) – (5.20)

$$c^* = 2.78330 - \left(\frac{0.06611}{2.78330} + \frac{1.61250}{93.42635}\right) +$$

$$\frac{1}{2}\,[\,(-1.27420)\,(0.06611)$$

$$+ (0.03480)\,(1.61250) + (0.19940)\,(1.42615)\,]$$

$$= 2.87041.$$

Similarly, $b^* = 96.22720$, $\mu* = 34.90182$.

At an arbitrary $t = 50$, from (5.22)

$$R^*_{50)} = 0.99500, \text{ the corresponding MLE}$$

$$\hat{R}_{50)} = 0.99631 \qquad \text{and the true}$$

$$R_{50)} = 0.99203.$$

Three-parameter lognormal

A three-parameter lognormal pdf of a random variable X is given by

$$f(x \mid k, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\,(x-k)\,\sigma} \exp\left[-\frac{1}{2\sigma^2}\{\log(x-k)-\mu\}^2\right],$$

$$x > k, -\infty < \mu < \infty, \sigma > 0. \tag{5.25}$$

Let the subscript (1, 2, 3) refer to the parameters (k, μ, σ) respectively, and let the prior $v(k, \mu, \sigma) \propto \frac{1}{\sigma^2}$.

We have $\quad \rho_1 = \rho_2 = 0, \ \rho_3 = -\frac{2}{\sigma}.$

Set the $\quad [\sigma_{ij}] = [-L_{ij}]^{-1}.$

$$\begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix} = -\begin{bmatrix} L_{11} & L_{12} & L_{13} \\ L_{21} & L_{22} & L_{23} \\ L_{31} & L_{32} & L_{33} \end{bmatrix}^{-1} \tag{5.26}$$

$[L_{ijk}]$ matrix given by (5.24) where

$$L = -\frac{n}{2}\log(2\pi) - n\log\sigma - \sum_{i=1}^{n}\log(x_i - k)$$

$$-\frac{1}{2\sigma^2}\sum_{i=1}^{n}[\log(x_i - k) - \mu]^2$$

$$L_{11} = \frac{\partial^2 L}{\partial k^2} = \sum_{i=1}^{n}\frac{1}{(x_i-k)^2}\left[1 + \frac{\log(x_i-k)-\mu-1}{\sigma^2}\right]$$

$$L_{22} = \frac{\partial^2 L}{\partial \mu^2} = -\frac{n}{\sigma^2}$$

$$L_{33} = \frac{\partial^2 L}{\partial \sigma^2} = \frac{n}{\sigma^2} - \frac{3}{\sigma^4} \sum_{i=1}^{n} [\log (x_i - k) - \mu]^2$$

$$L_{12} = \frac{\partial^2 L}{\partial k\, \partial \mu} = -\frac{1}{\sigma^2} \sum_{i=1}^{n} \frac{1}{(x_i - k)}$$

$$L_{13} = \frac{\partial^2 L}{\partial k\, \partial \sigma} = -\frac{2}{\sigma^3} \sum_{i=1}^{n} \left[\frac{\log (x_i - k) - \mu}{x_i - k}\right]$$

$$L_{23} = \frac{\partial^2 L}{\partial \mu\, \partial \sigma} = -\frac{2}{\sigma^3} \sum_{i=1}^{n} [\log (x_i - k) - \mu]$$

$$L_{111} = \frac{\partial^3 L}{\partial k^3} = \sum_{i=1}^{n} \frac{1}{(x_i - k)^3} \left[2 + \frac{2\{\log (x_i - k) - \mu\} - 3}{\sigma^2}\right]$$

$$L_{112} = \frac{\partial^3 L}{\partial k^2\, \partial \mu} = -\frac{1}{\sigma^2} \sum_{i=1}^{n} \frac{1}{(x_i - k)^2}$$

$$L_{113} = \frac{\partial^3 L}{\partial k^2\, \partial \mu} = \frac{2}{\sigma^3} \sum_{i=1}^{n} \left[\frac{1 - \log (x_i - k) + \mu}{(x_i - k)^2}\right]$$

$$L_{221} = \frac{\partial^3 L}{\partial \mu^2\, \partial k} = 0$$

$$L_{222} = \frac{\partial^3 L}{\partial \mu^3} = 0$$

$$L_{223} = \frac{\partial^3 L}{\partial \mu^2\, \partial \sigma} = \frac{2n}{\sigma^3}$$

$$L_{331} = \frac{\partial^3 L}{\partial \sigma^2\, \partial k} = \frac{6}{\sigma^4} \sum_{i=1}^{n} \left[\frac{\log (x_i - k) - \mu}{(x_i - k)}\right]$$

$$L_{332} = \frac{\partial^3 L}{\partial \sigma^2\, \partial \mu} = \frac{6}{\sigma^4} \sum_{i=1}^{n} [\log (x_i - k) - \mu]$$

$$L_{333} = \frac{\partial^3 L}{\partial \sigma^3} = \frac{-2n}{\sigma^3} + \frac{12}{\sigma^5} \sum_{i=1}^{n} [\log(x_i - k) - \mu]^2$$

all functions to be evaluated at the MLE $\hat{\theta} = (\hat{k}, \hat{\mu}, \hat{\sigma})$.

Substituting in (5.18) – (5.20) and replacing (c, b, μ) by (k, μ, σ) respectively, we obtain $\theta^* = (k^*, \mu^*, \sigma^*)$.

Lye, Sinha and Booy (1988) proposed a Bayesian analysis of flood data fitted by a three-parameter lognormal distribution.

Example 5.2:

A random sample of size $n = 50$ was generated from the lognormal pdf (5.25) with $k = 6$, $\mu = -1$, $\sigma = \frac{1}{2}$. The MLE $\hat{\theta}$ was obtained as $\hat{\theta} = (\hat{k}, \hat{\mu}, \hat{\sigma}) = (6.096, -1.467, 0.811)$. We will compute Bayes estimator $\theta* = (k^*, \mu^*, \sigma^*)$. At $\hat{\theta}$, we have

$$L_{11} = -8553.1411, \quad L_{12} = -458.9951,$$

$$L_{13} = 748.6348, \quad L_{22} = -76.0201,$$

$$L_{23} = 0.2088, \quad L_{33} = -152.5644,$$

$$\rho_1 = \rho_2 = 0, \quad \rho_3 = -2.4661.$$

Substituting in (5.26)

$$\begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix} = \begin{bmatrix} 0.0005 & -0.0028 & 0.0023 \\ -0.0028 & 0.0303 & -0.0139 \\ 0.0023 & -0.0139 & 0.0179 \end{bmatrix}.$$

The elements of the L_{ijk} matrix (5.24) at $\hat{\theta} = (\hat{k}, \hat{\mu}, \hat{\sigma})$ are given by

$$L_{111} = -518519.5669, \; L_{112} = -5293.3856, \; L_{113} = 29678.6980,$$

$$L_{221} = 0, \; L_{222} = 0, \; L_{223} = 187.4724, \; L_{331} = -2769.3028,$$

$$L_{332} = -0.7725, \; L_{333} = 939.9481.$$

We may now compute, A, B, D and substituting in (5.18) – (5.20) we obtain $K^* = 6.0765$, $\mu^* = -1.3656$, $\sigma^* = 0.7557$.

Bulk Sampling Problem

Standard quality control techniques deal with the inspection of discrete units of products. A random sample of items is drawn from the lot and one or more quality characteristic(s) is measured on each item in the sample. The measurement may be an attribute (defective or non-defective) or a variable (length, time to failure, etc.). The information collected from the sample is used to construct the appropriate quality control charts or acceptance sampling plans.

However, the product may consist of material in the bulk form, such a truck-load of cement, coal, milk or wool. Sampling from such products is known as bulk sampling.

Hapuarachchi and Sinha (1990) proposed a Bayesian approach to construct an appropriate sampling plan for bulk material.

Suppose a random sample of n_1 primary units, (such as n_1 bales from a shipment of wool) is drawn from a lot containing N primary units and n_2 measurements are made on each primary unit. Let X_{ij} be the j^{th} measurement of i^{th} primary unit. We may model X_{ij} as

$$X_{ij} = \mu + \alpha_i + e_{ij}, \tag{5.27}$$

where,

μ = general process or lot mean,

α_i = between batches variability,

e_{ij} = random errors, $i = 1, 2, \ldots n_1$; $j = 1, 2, \ldots, n_2$.

We assume that e_{ij} and α_i are independently normally distributed with zero means and variances σ_1^2 and σ_2^2, respectively.

The acceptance sampling plans for the process mean μ are usually derived using the sample mean, $\overline{X} = \sum \sum X_{ij} / n$. It is easy to show that

$$\operatorname{Var}(\overline{X}) = (1 - n_1/N)\, \sigma_2^2/n_1 - \frac{\sigma_1^2}{n} = \frac{\sigma_2^2}{n_1} + \frac{\sigma_1^2}{n},\ n = n_1 n_2, \text{ if } N \text{ is large}$$

(Cochran, 1963).

The variances σ_1^2 and σ_2^2 are generally unknown and have to be estimated in order to construct an acceptance sampling plan. Quality control engineers and practitioners usually use the components of variation estimates $(\tilde{\sigma}_1^2, \tilde{\sigma}_2^2)$ from the analysis of variance table. Let,

$$\bar{X}_i = \sum_j X_{ij}/n_2,$$

$$S_1 = \sum_i \sum_j (X_{ij} - \bar{X}_i)^2,$$

$$S_2 = n_2 \sum_i (\bar{X}_i - \bar{X})^2,$$

$$\nu_1 = n_1(n_2-1),\ \nu_2 = n_1 - 1,\ m_1 = S_1/\nu_1,\ m_2 = S_2/\nu_2.$$

The analysis of variance table for the model (5.27) is given by

Anova Table

Sources	*Df*	*SS*	*MSS*	*E (MSS)*
Between Sampling	ν_2	S_2	$m_2 = S_2/\nu_2$	$\sigma_1^2 + \sigma_1^2 + n_2\sigma_2^2$
Within Sampling units (Error)	ν_1	S_1	$m_1 = S_1/\nu_1$	σ_1^2.
Total	$n_1 n_2 - 1$	$\sum_i \sum_j (X_{ij} - \bar{X})^2$	(Box and Tiao, 1973)	

Thus, $\tilde{\sigma}_1^2 = m_1,\ \tilde{\sigma}_2^2 = (m_2 - m_1)/n_2.$

$\tilde{\sigma}_1^2$ and $\tilde{\sigma}_2^2$ are unbiased. However, if $m_1 > m_2$, $\tilde{\sigma}_2^2 < 0$ which is unacceptable. To avoid this possibility we obtain Bayes estimates of $\tilde{\sigma}_1^2$ and $\tilde{\sigma}_2^2$.

From (5.27) it follows that X_{ij}'s are normally distributed with mean μ and variance $\tilde{\sigma}_1^2 + \tilde{\sigma}_2^2$. For brevity let us write $\sigma_1^2 = r_1$, $\sigma_2^2 = r_2$. Let $\underline{x} = (x_1, x_2, \cdots, x_n)$ be the data, $\theta = (r_1, r_2, \mu)$ and using Jeffreys (1961) principle of ignorance prior, let the joint prior for (r_1, r_2, μ) be given by

$$v\,(r_1, r_2, \mu) \propto \frac{1}{r_1\, r_2}.$$

After some algebra one obtains the likelihood function $l\,(r_1, r_2, \mu \mid x) \propto$

$$r_1^{-\nu/2}\left(r_1 + n_2 r_2\right)^{-\left(\frac{\nu_2}{2}+1\right)} \exp\left[-\frac{1}{2}\left\{\frac{\nu_1 m_1}{r_1} + \frac{\nu_2 m_2 + n_1 n_2 (\bar{x} - \mu)^2}{r_1 + n_2 r_2}\right\}\right].$$

Combining the prior with the likelihood, we obtain the joint posterior distribution of (r_1, r_2, μ):

$$\Pi(r_1, r_2, \mu \mid \underline{x}) \propto r_1^{-(v_1/2+1)}\; r_2^{-1}\,(r_1 + n_2 r_2)^{-(v_2/2+1)}$$

$$\exp\left[\frac{1}{2}\left(\frac{v_1 m_1}{r_1} + \frac{v_2 m_2 + n_1 n_2 (\bar{x} - \mu)^2}{r_1 + n_2 r_2}\right)\right].$$

We integrate out (r_1, r_2, μ) in turn and derive the corresponding marginal posteriors. Under squared-error loss function, Bayes' estimator of θ is the posterior expectation of θ and the posterior expectations of (r_1, r_2, μ) involve complicated double integrals which cannot be expressed in closed forms. We use, instead, Lindley's (1980) approximation (5.17), which does not require special tables or iterative process and is much easier to work with.

For the prior (5.28), $\rho_1 = -1/r_1$, $\rho_2 = -1/r_2$, $\rho_3 = 0$. To obtain the Bayes estimator of r_1, put $u = r_1$, $u_1 = 1$, $u_2 = u_3 = 0$, $u_{ij} = 0$, $i,j = 1, 2, 3$, $a_4 = a_5 = 0$, $a_1 = -(\sigma_{11}/r_1 + \sigma_{12}/r_2)$.

Substituting in (5.17)

$$r_1^* = E(r_1/\underline{x}) = [r_1 - (\sigma_{11}/r_1 + \sigma_{12}/r_2)$$
$$+ 1/2\,(A\sigma_{11} + B\sigma_{12} + C\sigma_{31})]_{\hat{\theta}}.$$

Similarly,

$$r_2^* = [r_2 - (\sigma_{21}/r_1 + \sigma_{22}/r_2) + 1/2\,(A\sigma_{21} + B\sigma_{22} + C\sigma_{23})]_{\hat{\theta}}$$

$$\mu^* = [\mu - (\sigma_{31}/r_1 + \sigma_{32}/r_2) + 1/2\,(A\sigma_{31} + B\sigma_{32} + C\sigma_{33})]_{\hat{\theta}} \qquad (5.29)$$

where $\hat{\theta} = (\hat{r}_1, \hat{r}_2, \hat{\mu})$.

Again, all functions are to be evaluated at the maximum likelihood estimates

$$\hat{\mu} = \bar{x}$$

$$\hat{r}_1 = m_1$$

$$\hat{r}_2 = \left[\left(\frac{n_1 - 1}{n_1}\right) m_2 - m_1\right] \Big/ n_1 \quad \text{if} \quad \left(\frac{n_1 - 1}{n_1}\right) m_2 > m_1$$

and $\hat{r}_1 = (\nu_1 m_1 + \nu_2 m_2) / n_1 n_2,$ if $\left(\frac{n_1 - 1}{n_1}\right) m_2 < m_1$

$\hat{r}_2 = 0.$ (Herbach, 1959).

If $\left(\frac{n_1 - 1}{n_1}\right) m_2 < m_1$ and we accept $\hat{r}_2 = 0$ as the maximum likelihood estimator of r_2, then $\hat{\rho}_2 = \frac{1}{\hat{r}_2} = \infty$ and Lindley's expansion will not work. In such a case, one may use Jeffreys' invariant prior,

$$g(r_1, r_2) \propto \frac{1}{r_1 (r_2 + n_2 r_2)} \quad \text{(Box and Tiao, 1973)} \qquad (5.30)$$

and the corresponding Bayes estimators are given by

$$r_1^* = \left[r_1 - \left\{ \frac{\sigma_{11}(2r_1 + n_2 r_2) + n_2 \sigma_{12} r_1}{r_1 (r_1 + n_2 r_2)} \right\} + 1/2 \, (A\sigma_{11} + B\sigma_{12} + C\sigma_{13}) \right]_\theta$$

$$r_2^* = \left[r_2 - \left\{ \frac{\sigma_{21}(2r_1 + n_2 r_2) + n_2 \sigma_{22} r_1}{r_1 (r_1 + n_2 r_2)} \right\} + 1/2 \, (A\sigma_{21} + B\sigma_{22} + C\sigma_{23}) \right]_\theta$$

$$\mu^* = \left[\mu - \left\{ \frac{\sigma_{31}(2r_1 + n_2 r_2) + n_2 \sigma_{32} r_1}{r_1 (r_1 + n_2 r_2)} \right\} + 1/2 \, (A\sigma_{31} + B\sigma_{32} + C\sigma_{33}) \right]_\theta$$

Computations of posterior variances of (r_1, r_2, μ) under priors (5.28) and (5.30) are left as exercises for the readers.

Example 5.3:

(i) In order to estimate the average percent potassium bitartrate content in a shipment of argol, ten argol bags were randomly selected and two measurements of percent potassium bitartrate content from each bag were obtained (Tanner and Lerner, 1951). In this case, σ_1^2 represents the "within bag" variability and σ_2^2 is the "between bag" variability. The following is the summary of various computations

$$m_1 = .3124, \quad m_2 = .7385, \quad \bar{x} = 89.292,$$

$$n_1 = 10. \quad n_2 = 2, \quad \nu_1 = 10, \quad \nu_2 = 9.$$

$$\sigma_{ij} = \begin{bmatrix} .0195188 & -.0097594 & 0 \\ -.009594 & .0269677 & 0 \\ 0 & 0 & 0.332325 \end{bmatrix}$$

$$\begin{aligned}
&L_{111} = 724.10584, \quad L_{222} = 544.9303, \\
&L_{333} = 0, \quad L_{112} = 136.23257, \\
&L_{221} = 272.46515, \quad L_{331} = 45.27349, \\
&L_{113} = 0, \quad L_{223} = 0, \quad L_{332} = 136.23257, \\
&A = 20.32689, \ B = 16.56379, \ C = 0.
\end{aligned}$$

The maximum likelihood estimates are $\hat{r}_1 = .3124$, $\hat{r}_2 = .176125$, $\hat{\mu} = 89.292$.

The Bayes estimators and their posterior variances have been computed and the results are summarized in Table 1. In this table, method of estimation (1) and (2) refers to the prior distributions (5.28) and (5.30) respectively. The estimates of variances are given in the parentheses.

Table I Estimates of σ_1^2 and σ_2^2

	Method of Estimation			
			Bayes	
Parameter	*MLE*	*Moment*	(1)	(2)
σ_1^2	.3124	.3124	.4229	.3675
	(.0195)	(.0195)	(.0073)	(.0165)
σ_2^2	.1761	.2131	.1784	.2651
	(.0269)	(.0352)	(.0268)	(.0191)

Table I indicates that the posterior variances of σ_1^2 and σ_2^2 are less than their corresponding maximum likelihood counterpart.

Now consider constructing an appropriate sampling plan for the acceptance/rejection of the shipment of argol. For this purpose we use posterior variance estimates of σ_1^2 and σ_2^2 based on the prior distribution (5.28) to estimate the variance of the sample mean. Thus,

$$\operatorname{Var}(\bar{X}) = \frac{.0268}{10} + \frac{.0073}{10(2)}$$

$$= .00268 + .000365 = .003045.$$

The estimated standard deviation of sample mean $= \sqrt{.003045} = .055182$.

Now assume that the shipment of argol should contain at most 90 percent of potassium bitartrate (i.e., the upper specification limit is 90 percent). The value of the test statistic is $Z = (90 - 89.292)/(.055182) = 12.83$. Let producer's risk α be .05. Since $12.83 > 1.645$ the shipment of argol is considered unacceptable.

(ii) In this example we illustrate the estimation of the components of variance in (5.27) when $m_2 < m_1$. Box and Tiao (1973, pp. 247) generated 30 observations from the model (5.27) such that

$n_1 = 6$, $n_2 = 5$, $\sigma_1^2 = 16$, $\sigma_2^2 = 4$. The various computations associated with this example are:

$$m_1 = 14.9459,\ m_2 = 8.3363,\ \bar{x} = 5.6656,\ \nu_1 = 24,\ \nu_2 = 5.$$

$$\sigma_{ij} = \begin{bmatrix} 11.972876 & -2.3945876 & 0 \\ -2.3945876 & 5.8346566 & 0 \\ 0 & 0 & .448588 \end{bmatrix}$$

$L_{111} = .0414905$, $L_{222} = .1771732$, $L_{333} = .0$, $L_{112} = .0070869$,
$L_{221} = .0354347$, $L_{331} = .1684273$, $L_{113} = .0$, $L_{223} = .0$,
$L_{332} = .0630998$, $L_{123} = .0$.

$A = .7444959$, $B = .9769629$, $C = 0$. Note that in this case $m_2 < m_1$ and thus, the maximum likelihood estimates of σ_1^2 and σ_2^2 are 13.3461 and 0, respectively.

As indicated earlier, Lindley's expansion does not work for the prior distribution in (5.28) and as such, Bayes estimators of σ_1^2 and σ_2^2 are obtained by using Jeffreys' invariant prior in (5.30). The results are given in Table II. Once again, the corresponding estimated variances of variance components are given in the parentheses.

Table II Estimates of σ_1^2 and σ_2^2

Parameter	*True*	*MLE*	*Bayes*
σ_1^2	16	13.3461 (11.9729)	15.7362 (6.2605)
σ_2^2	4	0 (5.8347)	.1317 (5.8173)

We note that the posterior variances are less than their corresponding maximum likelihood counterpart.

5.3 TIERNEY–KADANE'S (T–K) APPROXIMATION

Lindley's approximation (5.2) is accurate enough but one of the problems of this approximation is that it requires evaluation of third order partial derivatives and in a p-parameter case, the total number of such derivatives is $\frac{p\,(p+1)\ (p+2)}{6}$. For the 3-parameter Weibull distribution we need 10 third order derivates, L_{111}, L_{222}, L_{333}, L_{223}, L_{112}, L_{221}, L_{132}, L_{113}, L_{331} and L_{332}.

Tierney-Kadane (1986) introduced an approximation for the posterior expectation,

$$E\,\{u\,(\theta)\,|\,\underline{x}\} = \frac{\int u(\theta)\exp\,\{L(\theta)\}\,v(\theta)\,d\theta}{\int v(\theta)\exp\,\{L(\theta)\}\,d\theta} \tag{5.31}$$

using Laplace's method for integrals (DeBruijn, 1961) which requires only first and second derivatives.

We may re-write (5.31) as

$$E\,\{u(\theta)\,|\,\underline{x}\} = \frac{\int \exp\,\{L(\theta) + \log v(\theta) + \log u(\theta)\}\ d\theta}{\int \exp\,\{L(\theta) + \log v(\theta)\}\ d\theta}$$

$$= \frac{\int \exp\,\{nL_*(\theta)\}\ d\theta}{\int \exp\,\{nL_0(\theta)\}\ d\theta} \tag{5.32}$$

$$L_0(\theta) = \frac{1}{n}\,[L(\theta) + \log v(\theta)]$$

where,

$$L_*(\theta) = L_0(\theta) + \frac{1}{n}\log u(\theta)$$

and $u(\theta)$ is an arbitrary positive function of θ.

Setting $\left[\frac{\partial^2 L_0}{\partial\theta^2}\right]_{\theta_0} = -\frac{1}{\sigma_0^2}$, for large n,

$$= \int \exp\{nL_0(\theta)\}\, d\theta$$

$$= \int \exp\left\{ nL_0(\theta_0) + \frac{n}{2}(\theta-\theta_0)^2 \left.\frac{\partial^2 L_0}{\partial\theta^2}\right|_{\theta_0} \right\} d\theta$$

$$= \exp\{nL_0(\theta_0)\} \int \exp\left\{-\frac{n(\theta-\theta_0)^2}{2\sigma_0^2}\right\} d\theta$$

$$= \exp\{nL_0(\theta_0)\} \sqrt{\frac{2\pi}{n}}\, \sigma_0.$$

Similarly, $\int \exp\{nL_*(\theta)\}\, d\theta$

$$= \exp\{nL_*(\theta_*)\} \sqrt{\frac{2\pi}{n}}\, \sigma_*.$$

where $L_0(\theta_0) = \max\{L(\theta)\}_{\theta_0}$

and $L_*(\theta_*) = \max\{L_*(\theta)\}_{\theta_0}$.

Substituting in (5.32), we have the T–K approximation

$$u^*_{(\theta)} = E\{u(\theta) \mid \underline{x}\}$$

$$= \frac{\sigma_*}{\sigma_0} \exp[n\{L_*(\theta_*) - L_0(\theta_0)\}]. \tag{5.33}$$

For a multiparameter case, we have

$$u^*_{(\theta)} = \left[\frac{\det \sum_*}{\det \sum_0}\right]^{\frac{1}{2}} \exp[n\{L_*(\underline{\theta}_*) - L_0(\underline{\theta}_0)\}] \tag{5.34}$$

Tierney and Kadane (1986) refers the formula (5.33) as Laplace-approximation.

Although T–K approximation requires only the first and second derivative of L, the numerator in (5.32) has to be maximized separately for each function $u(\theta)$ to obtain θ_*. Computationally, it is not difficult but analytically it is considerably more tedious and much less appealing than Lindley's approximation (5.3). We will apply T–K approximation to the following distributions.

Exponential

$$f(x \mid \theta) = \frac{1}{\theta} \exp\left(-\frac{x}{\theta}\right), \quad x, \theta > 0.$$

Let $\quad v(\theta) \propto \dfrac{1}{\theta^2}$

$$nL_0 = -n \log \theta - \frac{S}{\theta} - 2 \log \theta,$$

$$= -(n+2) \log \theta - \frac{S}{\theta}, \quad S = \sum_{i=1}^{n} x_i .$$

$$\frac{\partial L_0}{\partial \theta} = 0 = -\frac{n+2}{\theta} + \frac{S}{\theta^2}. \text{ Hence,}$$

$$\theta_0 = \frac{S}{n+2}$$

$$n\frac{\partial^2 L_0}{\partial \theta^2} = \frac{(n+2)}{\theta^2} - \frac{2S}{\theta^3}$$

$$-n\frac{\partial^2 L}{\partial \theta^2} = \frac{-(n+2)}{\theta_0^2} + \frac{2(n+2)}{\theta_0^2} = \frac{(n+2)}{\theta_0^2}.$$

$$\sigma_0^2 = \frac{n\theta_0^2}{n+2}.$$

Let $\quad u(\theta) = \theta.$

$$nL_* = \left[-(n+2) \log \theta - \frac{S}{\theta} + \log \theta\right]_{\theta_*}$$

$$= \left[-(n+1)\log\theta - \frac{S}{\theta} \right]_{\theta_*}$$

$$\left[\frac{\partial L_*}{\partial \theta} \right]_{\theta_*} = 0 = \left[-\frac{n+1}{\theta} + \frac{S}{\theta^2} \right]_{\theta_*}.$$

Hence, $\theta_* = \dfrac{S}{n+1}$.

$$\left[n\frac{\partial^2 L_*}{\partial \theta^2} \right]_{\theta_*} = \left[\frac{n+1}{\theta^2} - \frac{2S}{\theta^3} \right]_{\theta_*}$$

$$= \left[\frac{n+1}{\theta^2} - \frac{2(n+1)}{\theta^2} \right]_{\theta_*} = \frac{(n+1)}{(\theta_*)^2}$$

$$(\sigma^*)^2 = \frac{n\theta_*^2}{n+1}$$

$$\frac{\theta^*}{\sigma_0} = \sqrt{\frac{n+2}{n+1}} \left(\frac{\theta_*}{\theta_0} \right) = \left(\frac{n+2}{n+1} \right)^{3/2}. \tag{5.35}$$

$$n(L^* - L_0) = (n+2)\log\theta_0 - (n+1)\log\theta_* + S\left(\frac{1}{\theta_0} - \frac{1}{\theta_*} \right)$$

$$= (n+1)\log\left(\frac{\theta_0}{\theta_*} \right) + \log\theta_0 + 1$$

$$\exp[n(L^* - L_0)] = \theta_0 \left(\frac{\theta_0}{\theta_*} \right)^{n+1} \cdot e \tag{5.36}$$

From (5.35) and (5.36), we obtain

$$E(\theta \mid \underline{x}) = \left(\frac{n+2}{n+1} \right)^{3/2} \left(\frac{n+1}{n+2} \right)^{n+1} \cdot e \cdot \left(\frac{S}{n+2} \right)$$

$$\to \bar{x} \text{ as } n \to \infty$$

which is the same as the one we have obtained in (5.5) using Lindley's approximation with much less algebra

Inverse Gaussian

Consider the $IG\ (\lambda, \mu)$ pdf (5.14).

$$L = K + \frac{n}{2}\log\lambda - \frac{\lambda}{2}\left[\frac{n\bar{x}}{\mu^2} - \frac{2n}{\mu} + \sum_i \frac{1}{x_i}\right],\quad i = 1, 2, \cdots, n.$$

$$\nu(\lambda, \mu) \propto \frac{1}{\lambda}.$$

For $u(\theta) = \lambda$,

$$L_0 = \frac{1}{n}\left[K + \left(\frac{n}{2} - 1\right)\log\lambda - \frac{\lambda}{2}\left\{\frac{n\bar{x}}{\mu^2} - \frac{2n}{\mu} + \sum_i \frac{1}{x_i}\right\}\right]$$

$$L_* = \frac{1}{n}\left[K + \frac{n}{2}\log\lambda - \frac{\lambda}{2}\left\{\frac{n\bar{x}}{\mu^2} - \frac{2n}{\mu} + \sum_i \frac{1}{x_i}\right\}\right]$$

$$\frac{\partial L_0}{\partial\lambda} = \frac{1}{n}\left[\left(\frac{n}{2} - 1\right)\frac{1}{\lambda} - \frac{1}{2}\left\{\frac{n\bar{x}}{\mu^2} - \frac{2n}{\mu} + \sum_i \frac{1}{x_i}\right\}\right]$$

$$\frac{\partial L_0}{\partial\mu} = \frac{1}{n}\left[-\frac{\lambda}{2}\left\{\frac{-2n\bar{x}}{\mu^3} + \frac{2n}{\mu^2}\right\}\right].$$

$$\frac{\partial L_0}{\partial\mu} = 0 \text{ yields } \mu_0 = \bar{x} \text{ and}$$

$$\frac{\partial L_0}{\partial\lambda} = 0 \text{ implies } \left(\frac{n}{2} - 1\right) = \frac{\lambda_0}{2}\left\{\frac{n}{\bar{x}} - \frac{2n}{\bar{x}} + \sum_i \frac{1}{x_i}\right\}.$$

Thus, $$(\mu_0, \lambda_0) \equiv \left[\bar{x}, \frac{n-2}{\sum_i\left(\frac{1}{x_i} - \frac{1}{\bar{x}}\right)}\right] \equiv \left[\hat{\mu}, \left(\frac{n-2}{n}\right)\hat{\lambda}\right].$$

Similarly $$\frac{\partial L_*}{\partial\mu} = 0 \text{ and}$$

$$\frac{\partial L_*}{\partial\lambda} = 0 \text{ yield}$$

$$(\mu^*, \lambda^*) \equiv \left[\bar{x}, \frac{n}{\sum_i \left(\frac{1}{x_i} - \frac{1}{\bar{x}}\right)}\right] \equiv [\hat{\mu}, \hat{\lambda}].$$

$$n(L^* - L_0) = \frac{n}{2}\log\left(\frac{\lambda^*}{\lambda_0}\right) + \log\lambda_0 - \frac{1}{2}(\lambda^* - \lambda_0)\sum_i\left(\frac{1}{x_i} - \frac{1}{\bar{x}}\right)$$

$$= \log\left[\left(\frac{\lambda^*}{\lambda_0}\right)^{n/2} \cdot \lambda_0\right] - \frac{\lambda^*}{n}\sum_i\left(\frac{1}{x_i} - \frac{1}{\bar{x}}\right)$$

$$= \log\left[\left(\frac{\lambda^*}{\lambda_0}\right)^{n/2} \cdot \lambda_0\right] - 1$$

$$\exp[n(L_* - L_0)] = \exp(-1)\left[\lambda_0\left(\frac{\lambda^*}{\lambda_0}\right)^{n/2}\right]. \tag{5.37}$$

$$\frac{\partial^2 L}{\partial\lambda^2} = -\frac{1}{\lambda^2}\left(\frac{1}{2} - \frac{1}{n}\right)$$

$$\frac{\partial^2 L}{\partial\lambda\,\partial\mu} = \frac{\bar{x}}{\mu^2} - \frac{1}{\mu^2}$$

$$\frac{\partial^2 L}{\partial\mu^2} = -\frac{\lambda}{2n}\left(\frac{6n\bar{x}}{\mu^4} - \frac{4n}{\mu^3}\right).$$

At (μ_0, λ_0) we have

$$\Sigma_0 = \begin{bmatrix} \frac{\lambda_0}{(\bar{x})^3} & 0 \\ 0 & \frac{n-2}{2n\lambda_0^2} \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} \frac{(\bar{x})^3}{\lambda_0} & 0 \\ 0 & \frac{2n\lambda_0^2}{n-2} \end{bmatrix}.$$

$$|\Sigma_0| = \frac{2n\lambda_0(\bar{x})^3}{n-2}.$$

Similarly, at (μ^*, λ^*) we have

$$\left|\Sigma^*\right| = 2(\bar{x})^3 \lambda^* .$$

$$\frac{\left|\Sigma^*\right|}{\left|\Sigma_0\right|} = \frac{\lambda^*(n-2)}{n\lambda_0} = 1.$$

From (5.34) and (5.37) we have

$$\lambda^*_{T-K} = \lambda_0 \left(\frac{\lambda^*}{\lambda_0}\right)^{n/2} \cdot \exp(-1)$$

$$= \left(\frac{n-2}{2}\right)\left(\frac{n-2}{n}\right)^{n/2} \hat{\lambda} \exp(-1)$$

$$\to \hat{\lambda} \text{ as } n \to \infty .$$

For $\mu(\theta) = \mu$, we have

$$(\mu_0, \lambda_0) = \left[\hat{\mu}, \left(\frac{n-2}{2}\right)\hat{\lambda}\right].$$

$$L_* = \frac{1}{n}\left[K + \left(\frac{n}{2} - 1\right)\log\lambda - \frac{\lambda}{2}\left\{\frac{n\bar{x}}{\mu^2} - \frac{2n}{\mu} + \sum_i \frac{1}{x_i}\right\} + \log\mu\right]$$

$$\frac{\partial L_*}{\partial \mu} = \frac{1}{n\mu} + \frac{\lambda}{\mu^3}(\bar{x} - \mu)$$

$$\frac{\partial L_*}{\partial \lambda} = \frac{1}{n}\left[\left(\frac{n}{2} - 1\right)\frac{1}{\lambda} - \frac{1}{2}\left\{\frac{n\bar{x}}{\mu^2} - \frac{2n}{\mu} + \sum_i \frac{1}{x_i}\right\}\right]$$

and (λ^*, μ^*) are solutions of the simultaneous equations

$$(\bar{x} - \mu)\, n\lambda + \mu^2 = 0$$

$$\frac{n-2}{\lambda} = \frac{n\bar{x}}{\mu^2} - \frac{2n}{\mu} + \sum_i \frac{1}{x_i} .$$

No closed form solution will lead to a simple form for μ^*_{T-K} and a computer maximization routine such as Newton-Raphson has to be used.

Using Chhikara and Folks (1977) data consisting of 46 repair times for an airborne communication system, we obtain the MLE, T-K and Lindley estimator of λ and μ.

Parameter	*MLE*	*T–K*	*Lindley*
λ	1.6589	1.6216	1.6228
μ	2.6064	4.0390	4.1178

Normal Distribution N (μ, σ^2)

$$f(x+\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\,\sigma}\exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right], \quad -\infty < x, \mu < \infty,\ \sigma > 0.$$

Let $\quad v(\mu,\sigma) \propto \frac{1}{\sigma},\ u(\theta) = \sigma.$

$$L = K - n\log\sigma - \frac{1}{2\sigma^2}\sum_i (x_i-\mu)^2, \quad i = 1, 2, \cdots, n.$$

$$L_0 = \frac{1}{n}\left[K-(n+1)\log\sigma - \frac{1}{2\sigma^2}\sum_i (x_i-\mu)^2\right]$$

$$L_* = \frac{1}{n}\left[K - n\log\sigma - \frac{1}{2\sigma^2}\sum_i (x_i-\mu)^2\right]$$

$$\frac{\partial L_0}{\partial\mu} = \frac{\sum_i (x_i-\mu)}{n\sigma^2} = \frac{\bar{x}-\mu}{\sigma^2}$$

$$\frac{\partial L_0}{\partial\sigma} = \frac{1}{n}\left[-\frac{n+1}{\sigma} + \frac{\sum_i (x_i-\mu)^2}{\sigma^3}\right]$$

$$\frac{\partial L_0}{\partial\mu} = 0 \text{ yields } \mu_0 = \bar{x}$$

$$\frac{\partial L_0}{\partial\sigma} = 0 \text{ implies } \sigma_0^2 = \frac{\sum_i (x_i-\bar{x})^2}{n+1} = \frac{S^2}{n+1}.$$

We have $(\mu_0, \sigma_0) = \left[\bar{x}, \sqrt{\frac{S^2}{n+1}}\right]$ where $S^2 = \sum_i (x_i-\bar{x})^2$.

$$\frac{\partial L_*}{\partial \mu} = \frac{\bar{x} - \mu}{\sigma^2}$$

$$\frac{\partial L_*}{\partial \sigma} = \frac{1}{n}\left[-\frac{n}{\sigma} + \frac{\sum_i (x_i - \mu)^2}{\sigma^3}\right],$$

and we have $\quad (\mu_*, \sigma_*) = \left[\bar{x}, \sqrt{\frac{S^2}{n}}\right].$

$$\frac{\partial^2 L_0}{\partial \mu^2} = -\frac{1}{\sigma^2}$$

$$\frac{\partial^2 L_0}{\partial \mu\, \partial \sigma} = -\frac{2(\bar{x} - \mu)}{\sigma^3}$$

$$\frac{\partial^2 L_0}{\partial \sigma^2} = \frac{1}{n}\left[\frac{n+1}{\sigma^2} - \frac{3\,\Sigma\,(x_i - \mu)^2}{\sigma^4}\right]$$

$$= \frac{1}{n}\left[\frac{n+1}{\sigma_0^2} - \frac{3(n+1)\sigma_0^2}{\sigma_0^4}\right]_{(\mu_0,\, \sigma_0)}$$

$$= \frac{-2(n+1)}{n\sigma_0^2}$$

$$\sum_0 = \begin{bmatrix} \frac{1}{\sigma_0^2} & 0 \\ 0 & \frac{2(n+1)}{n\sigma_0^2} \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} \sigma_0^2 & 0 \\ 0 & \frac{n\sigma_0^2}{2(n+1)} \end{bmatrix}$$

$$\left|\sum_0\right| = \frac{n\sigma_0^4}{2(n+1)}$$

$$\frac{\partial^2 L_*}{\partial \mu^2} = -\frac{1}{\sigma^2}$$

$$\frac{\partial^2 L_*}{\partial\mu\,\partial\sigma} = -\frac{2(\bar{x}-\mu)}{\sigma^3}$$

$$\frac{\partial^2 L_*}{\partial\sigma^2} = -\frac{2}{\sigma_*^2} \text{ at } (\mu_*, \sigma_*)$$

$$\Sigma_* = \begin{bmatrix} \frac{1}{\sigma_*^2} & 0 \\ 0 & \frac{2}{\sigma_*^2} \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} \sigma_*^2 & 0 \\ 0 & \frac{\sigma_*^2}{2} \end{bmatrix}$$

$$\left|\Sigma_*\right| = \frac{\sigma_*^4}{2}$$

$$\left|\frac{\Sigma_*}{\Sigma_0}\right| = \frac{(n+1)\sigma_*^4}{n\sigma_0^4} = \left(\frac{n+1}{n}\right)\left(\frac{n+1}{n}\right)^2 = \left(\frac{n+1}{n}\right)^3$$

$$\left|\frac{\Sigma_*}{\Sigma_0}\right|^{1/2} = \left(\frac{n+1}{n}\right)^{3/2} \tag{5.38}$$

$$n(L_* - L_0) = \log\sigma_0 - n\log\left(\frac{\sigma_*}{\sigma_0}\right) - \frac{S^2}{2}\left(\frac{1}{\sigma_*^2} - \frac{1}{\sigma_0^2}\right)$$

$$= \log\sigma_0 - \frac{n}{2}\log\left(\frac{n+1}{n}\right) + \frac{S^2}{2(n+1)\sigma_0^2}$$

$$= \log\sigma_0 - \frac{n}{2}\log\left(\frac{n+1}{n}\right) + \frac{1}{2}$$

$$= \log\left\{\sigma_0\left(\frac{n+1}{n}\right)^{-n/2}\right\} + \frac{1}{2}$$

$$\exp\left[n\left(L_* - L_0\right)\right] = \frac{\sigma_0}{\left(\frac{n+1}{n}\right)^{n/2}} \exp\left(\frac{1}{2}\right)$$

$$\frac{\sqrt{\frac{n}{n+1}}\ \hat{\sigma}}{\left(\frac{n+1}{n}\right)^{n/2}} \exp\left(\frac{1}{2}\right). \tag{5.39}$$

From (5.38) and (5.39) we obtain

$$\sigma^*_{T-K} = \frac{\left(\frac{n+1}{n}\right)^{3/2} \cdot \left(\frac{n}{n+1}\right)^{1/2} \cdot \hat{\sigma} \exp\left(\frac{1}{2}\right)}{\left(\frac{n+1}{n}\right)^{n/2}}$$

$$= \frac{\left(\frac{n+1}{n}\right) \hat{\sigma} \exp\left(\frac{1}{2}\right)}{\left(\frac{n+1}{n}\right)^{n/2}}$$

$$= \hat{\sigma}\left(1 + \frac{1}{n}\right) \sqrt{\frac{e}{\left(1 + \frac{1}{n}\right)^n}}$$

$$\to \hat{\sigma}, \text{ the MLE of } \sigma, \text{ as } n \to \infty.$$

For $u(\theta) = \mu$, we will have the same $(\mu_0, \sigma_0) = \left[\bar{x},\ \sqrt{\frac{S^2}{n+1}}\right]$

$$L_* = \frac{1}{n}\left[K - (n+1)\log\sigma - \frac{1}{2\sigma^2}\sum_i (x_i - \mu)^2 + \log\mu\right]$$

$$\frac{\partial L_*}{\partial \mu} = \frac{1}{n\mu} + \frac{\bar{x} - \mu}{\sigma^2}$$

$$\frac{\partial L_*}{\partial \sigma} = -\frac{n+1}{n\sigma} + \frac{\sum_i (x_i - \mu)^2}{n\sigma^3}.$$

A closed form solution for (μ^*, σ^*) will be extremely messy and we have to use a computer maximization routine.

We note that Bayes estimators of (μ, σ) with the same prior $g(\mu, \sigma) \propto \frac{1}{\sigma}$ were obtained in (5.11) and (5.12) in simple closed forms using Lindley's approximation (Sloan & Sinha, 1990).

EXERCISES

1. Let $X \sim B(n, p)$ and prior $g(p) \propto \frac{1}{p(1-p)}$. Obtain Bayes estimator $E(p \mid x)$ using Lindley's approximation (5.3).
2. Let $X \sim N(\theta, 1)$, $\theta \sim N(\mu, 1)$. Derive $E(\theta \mid x)$ using (5.3).
3. Let $X \sim N(\mu, \sigma^2)$. Obtain Bayes estimator $E(\sigma \mid x)$ using (5.3) and the priors

   (i) $g(\mu, \sigma) \propto \frac{1}{\sigma^5}$ and

   (ii) $g(\mu, \sigma) \propto \frac{1}{\sigma}$.
4. Derive Lindley's approximation of $E(\theta \mid x)$ given

$$f(x \mid \theta) = \frac{1}{\theta} \exp\left(-\frac{x}{\theta}\right), \quad x > 0 \text{ and}$$

$$g(\theta) \propto \frac{1}{\theta}.$$

5. Derive Lindley's approximation of $E(\lambda \mid \underline{x})$, given $\underline{x} = (x_1, x_2, \cdots, x_n)$,

$$f(x \mid \lambda) = \frac{\lambda^x}{x!} \exp(-\lambda),\ \lambda > 0,\ x = 0, 1, 2, \cdots$$

and $g(\lambda) \propto \lambda^{b-1} \exp(-a\lambda),\ a, b > 0.$

6. Derive T–K estimator

   $E(\lambda \mid \underline{x})$ given $\underline{x} = (x_1, x_2, \cdots, x_n)$.

$$f(x \mid \lambda) = \frac{\lambda^x}{x!} \exp(-\lambda),\ \lambda > 0,\ x = 0, 1, 2, \cdots$$

$$g(\lambda) \propto \lambda^{b-1} \exp(-a\lambda), \quad a, b > 0.$$

7. Obtain T–K approximation of $E(p \mid x)$ given $X \sim B(n, p)$ and $g(p) \propto \dfrac{1}{p(1-p)}$.

8. Show that the T–K approximation $\rightarrow$ Lindley's approximation of $E(p \mid x)$ as $n \rightarrow \infty$.

9. Derive T–K approximation of $E(\theta \mid \underline{x})$, given $X \sim N(\theta, 1)$ and $\theta \sim N(\mu, 1)$.

10. Derive T–K approximation of $E(\sigma \mid \underline{x})$, given $X \sim N(0, \sigma^2)$ and $g(\sigma) \propto \dfrac{1}{\sigma}$.

11. Show that the posterior variances of r_1, r_2, r_3 in (5.29) are given by:

$$V(r_1 \mid \underline{x}) = \sigma_{11} - [1/2\,(A\sigma_{11} + B\sigma_{12} + C\sigma_{13}) - (\sigma_{11}/r_1 + \sigma_{12}/r_2)]^2$$

$$V(r_2 \mid \underline{x}) = \sigma_{22} - [1/2\,(A\sigma_{12} + B\sigma_{22} + C\sigma_{23}) - (\sigma_{12}/r_1 + \sigma_{22}/r_2)]^2$$

$$V(\mu \mid \underline{x}) = \sigma_{33} - [1/2\,(A\sigma_{31} + B\sigma_{32} + C\sigma_{33}) - (\sigma_{13}/r_1 + \sigma_{23}/r_2)]^2$$

12. Let $(x_1, x_2, \cdots, x_n)$ be a random sample from $N(\mu, \sigma^2)$. Compute Lindley's and T–K approximations of Bayes estimates (μ^*, σ^*) under the joint prior $v(\mu, \sigma) \propto \dfrac{1}{\sigma^5}$.

Data $\{x_i\}$, $n = 25$

1.3132	9.9619	10.5209	2.5789	7.7943
4.8404	5.3648	5.7808	7.0625	4.5249

4.8805	1.5292	2.1633	3.3948	8.3039
1.4677	10.2535	11.9162	7.2924	4.7166
9.8624	2.0173	7.0530	4.0189	8.4281

13. Let $(x_1, x_2, \cdots x_n)$ be a random sample from $IG(\lambda, \mu)$. Obtain Lindley's and T–K approximation of Bayes estimators (λ^*, μ^*) under the joint prior $\nu(\lambda, \mu) \propto \frac{1}{\lambda}$.

    Data $\{x_i\}$, $n = 20$

0.6533	2.3787	0.8363	5.8301	3.2766
2.2944	0.7878	4.6179	3.1661	1.3260
0.7369	4.4883	2.9131	1.0180	13.5246
3.9869	2.5890	0.8422	10.3276	3.3318

14. Dumonceaux and Antle (1973) cite data obtained in a civil engineering context of the maximum flood level (in millions of cubic feet per second) for the Susquehanna River at Harrisburg, Pennsylvania over 20 four-year periods as:

0.654	0.613	0.315	0.449	0.297
0.402	0.379	0.423	0.379	0.324
0.269	0.740	0.418	0.412	0.494
0.416	0.338	0.392	0.484	0.265

Consider the three-parameter $IG(\alpha, \lambda, \mu)$ model

$$f(x \mid \alpha, \lambda, \mu) = \left[\frac{\lambda}{2\pi(x-\mu)^3}\right]^{1/2} \cdot$$

$$\exp\left[-\frac{\lambda(x-\alpha-\mu)^2}{2\mu^2(x-\alpha)}\right]. \quad x > \alpha, \quad \lambda, \mu > 0$$

and the joint prior $\nu(\alpha, \lambda, \mu) \propto \frac{1}{\lambda}$.

The maximum likelihood estimates $(\hat{\alpha}, \hat{\mu}, \hat{\lambda}) \equiv (0.178, 0.245, 0.914)$ (Cheng and Amin, 1981).

Obtain Lindley's approximation (1980) of Bayes estimators of $(\alpha^*, \mu^*, \lambda^*)$.

CHAPTER

6 Bayesian Regression Analysis

6.1 Linear Regression

Let (x_1, y_1), (x_2, y_2), $\cdots$, (x_n, y_n), be n pairs of observation such that x's are non-random or fixed and y's are independent, identically distributed (i.i.d.) random variables. Suppose for each $x = x_i$, y_i is normally distributed with mean

$$E(Y_i \mid x_i) = \alpha + \beta(x_i - \bar{x}) \tag{6.1}$$

and variance σ^2.

We may write the model (6.1) as

$$Y_i = \alpha + \beta(x_i - \bar{x}) + \epsilon_i$$

where ϵ_i's are i.i.d. random errors ~ $N(0, \sigma^2)$.

Let $$U = \sum_{i=1}^{n} \epsilon_i^2 = \sum_i [y_i - \alpha - \beta(x_i - \bar{x})]^2.$$

The least square estimators (a, b) of (α, β) are solutions of the equations

$$\frac{\partial U}{\partial \alpha} = \sum_i [y_i - a - b(x_i - \bar{x})] = 0, \quad \text{and}$$

$$\frac{\partial U}{\partial \beta} = \sum_i (x_i - \bar{x})\,[y_i - a - b(x_i - \bar{x})] = 0.$$

We obtain $a = \bar{y}$, $\quad b = \dfrac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2} = \dfrac{S_{xy}}{S_{xx}}$.

The 'fitted' y_i is given by

$$\hat{y}_i = a + b(x_i - \bar{x})$$

$$= \bar{y} + b(x_i - \bar{x}).$$

Let S^2 = sum of squares of deviations of observed y's from fitted y's

$$= \sum_i (y_i - \hat{y}_i)^2$$

$$= \sum_i [y_i - \bar{y} - b(x_i - \bar{x})]^2$$

$$= S_{yy} + b^2 S_{xx} - 2b\,S_{xy}$$

$$= S_{yy} - b^2 S_{xx}$$

$$= S_{yy} - \frac{S_{xy}^2}{S_{xx}}.$$

$$\sum_i [y_i - \alpha - \beta(x_i - \bar{x})]^2$$

$$= \sum_i [(y_i - \bar{y}) + (\bar{y} - \alpha) - \beta(x_i - \bar{x})]^2$$

$$= S_{yy} + n(\alpha - \bar{y})^2 + \beta^2 S_{xx} - 2\beta S_{xy}$$

$$= S_{yy} + n(\alpha - a)^2 - \frac{S_{xy}^2}{S_{xx}} + S_{xx}\left[\beta^2 - 2\beta\frac{S_{xy}}{S_{xx}} + \left(\frac{S_{xy}}{S_{xx}}\right)^2\right]$$

$$= \left(S_{yy} - \frac{S_{xy}^2}{S_{xx}}\right) + n(\alpha - a)^2 + S_{xx}\left(\beta - \frac{S_{xy}}{S_{xx}}\right)^2$$

$$= S^2 + n(\alpha - a)^2 + S_{xx}(\beta - b)^2.$$

The likelihood function,

$$l(x, y \mid \alpha, \beta) \propto \frac{1}{(\sigma^2)^{n/2}} \exp\left[-\frac{1}{2\sigma^2} \sum_i \{(y-\alpha-\beta(x_i-\bar{x}))\}\right]^2$$

$$\propto \frac{1}{(\sigma^2)^{n/2}} \exp\left[-\frac{1}{2\sigma^2} \{S^2+n(\alpha-a)^2+S_{xx}(b-\beta)^2\}\right]. \tag{6.2}$$

6.2 Independent Prior

Suppose $(\alpha, \beta, \sigma^2)$ are *a-priori* independent and (α, β) are locally uniform. Let σ^2 be distributed as inverted gamma (Raiffa and Schlaifer, 1961), with parameters c and p. Let,

$$h(\sigma^2) = \frac{c^{p/2}}{\Gamma\left(\frac{P}{2}\right)} \frac{\exp\left(-\frac{c}{\sigma^2}\right)}{(\sigma^2)^{p/2+1}}.$$

Note that

$$E(\sigma^2) = \int_0^\infty \sigma^2 h(\sigma^2)\, d\sigma^2$$

$$= \frac{c^{p/2}}{\Gamma\left(\frac{p}{2}\right)} \int \frac{\exp\left(-\frac{c}{\sigma^2}\right)}{(\sigma^2)^{(p-2)/2+1}}\, d\sigma^2$$

$$= \frac{c\,\Gamma\left(\frac{p-2}{2}\right)}{\Gamma\left(\frac{p}{2}\right)}$$

$$= \frac{2c}{p-2}. \tag{6.3}$$

The joint prior of $(\alpha, \beta, \sigma^2)$ may be written as

$$g(\alpha, \beta, \sigma^2) \propto h(\sigma^2)$$

$$\propto \frac{\exp\left(-\dfrac{c}{\sigma^2}\right)}{(\sigma^2)^{p/2+1}}. \tag{6.4}$$

Let D represent the data $\{(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)\}$. Combining (6.2) and (6.4) we obtain the joint posterior

$$\Pi(\alpha, \beta, \sigma^2 \mid D) \propto \frac{1}{(\sigma^2)^{(n+p)/2+1}} \cdot$$

$$\exp\left[-\frac{1}{2\sigma^2\{S^2 + n(\alpha + a)^2 + 2c + S_{xx}(\beta - b)^2\}}\right] \tag{6.5}$$

Integrating out α and σ^2 from (6.5), we have the marginal posterior of β given by

$$\Pi_1(\beta \mid D) \propto \int_0^{\infty} \frac{1}{(\sigma^2)^{(n+p-1)/2+1}} \cdot$$

$$\exp\left[-\frac{1}{2\sigma^2}\{S^2 + 2c + S_{xx}(\beta - b)^2\}\right] d\sigma^2$$

$$\propto \frac{1}{[S^2 + 2c + S_{xx}(\beta - b)^2]^{(n+p-1)/2}}$$

$$\propto \frac{1}{\left[1 + \dfrac{(b-\beta)^2}{\left(\dfrac{S^2 + 2c}{S_{xx}}\right)}\right]^{(n+p-1)/2}}.$$

Thus,

$$\Pi_1(\beta \mid D) = \frac{K}{\left[1 + \dfrac{(\beta - b)^2}{\left(\dfrac{S^2 + 2c}{S_{xx}}\right)}\right]^{(n+p-1)/2}} \tag{6.6}$$

$$= \frac{K}{\left[1 + \dfrac{(\beta - b)^2}{A}\right]^{(\upsilon+1)/2}}$$

where $\upsilon = n + p - 2,\ A = \dfrac{S^2 + 2c}{S_{xx}}$.

$$K^{-1} = \int_{-\infty}^{\infty} \frac{d\beta}{\left[1 + \frac{(\beta - b)^2}{A}\right]^{(\upsilon+1)/2}}.$$

Let $$\frac{\beta - b}{\sqrt{A}} = \frac{t}{\sqrt{\upsilon}} \tag{6.7}$$

$$d\beta = \sqrt{\frac{A}{\upsilon}}\, dt$$

$$K^{-1} = \sqrt{\frac{A}{\upsilon}} \int_{-\infty}^{\infty} \frac{dt}{\left(1 + \frac{t^2}{\upsilon}\right)^{(\upsilon+1)/2}}$$

$$= \sqrt{\frac{A}{\upsilon}}\, \sqrt{\upsilon}\, B\left(\frac{1}{2}, \frac{\upsilon}{2}\right)$$

$$= \sqrt{A}\, B\left(\frac{1}{2}, \frac{\upsilon}{2}\right).$$

Substituting the normalizing constant K in (6.6) we have

$$\Pi_1(\beta \mid D) = \frac{1}{\sqrt{A}\, B\left(\frac{1}{2}, \frac{\upsilon}{2}\right)\left[1 + \frac{(\beta - b)^2}{A}\right]^{(\upsilon+1)/2}}. \tag{6.8}$$

From (6.7) it follows that $\sqrt{\frac{\upsilon}{A}}\,(\beta - b)$ is distributed as Student's-t with $(n + p - 2)$ degrees of freedom. We write

$$\sqrt{\frac{\upsilon}{A}}\,(\beta - b) \sim t(\upsilon).$$

Since $E(t) = 0$, Bayes estimator of β

$\beta^* = b$, the same as the least square estimator of β.

$$\sqrt{\frac{\upsilon}{A}}\,(\beta - b) = \frac{\beta - b}{\sqrt{\frac{S^2 + 2c}{(n + p - 2)\, S_{xx}}}} \sim t(n + p - 2).$$

Putting $c = p = 0$, we have independent uniform prior for α, β and $\log \sigma^2$

$$g(\alpha, \beta, \sigma^2) \propto \frac{1}{\sigma^2}. \tag{6.9}$$

(6.8) implies

$$\frac{\beta - b}{\sqrt{\dfrac{S^2}{(n-2)S_{xx}}}} \sim t(n-2). \qquad \text{(Lindley, 1965)}$$

Integrating out β and σ^2 from (6.5), the marginal posterior of α is given by

$$\Pi_2(\alpha/D) \propto \frac{1}{[S^2 + 2c + n(\alpha - a)^2]^{(n+p-1)/2}}$$

$$= \frac{K_1}{\left[1 + \dfrac{(\alpha - a)^2}{\left(\dfrac{S^2 + 2c}{n}\right)}\right]^{(n+p-1)/2}} \tag{6.10}$$

$$K_1^{-1} = \int_{-\infty}^{\infty} \frac{d\alpha}{\left[1 + \dfrac{(\alpha - a)^2}{\left(\dfrac{S^2 + 2c}{n}\right)}\right]^{(n+p-2+1)/2}}$$

$$= \sqrt{\frac{S^2 + 2c}{n(n+p-2)}} \int_{-\infty}^{\infty} \frac{dt}{\left[1 + \dfrac{t^2}{n+p-2}\right]^{(n+p-2+1)/2}}$$

$$= \sqrt{\frac{S^2 + 2c}{n}}\, B\left(\frac{1}{2}, \frac{n+p-2}{2}\right)$$

where $$\sqrt{\frac{n}{S^2 + 2c}}(\alpha - a) = \frac{t}{\sqrt{n+p-2}}. \tag{6.11}$$

Substituting K_1 in (6.10),

$$\Pi_2(\alpha \mid D) = \sqrt{\frac{n}{S^2 + 2c}}\,.$$

$$\left[\frac{1}{B\left(\frac{1}{2}, \frac{n+p-2}{2}\right)\left[1+\frac{(\alpha-a)^2}{\frac{S^2+2c}{n}}\right]^{\frac{(n+p-2+1)}{2}}}\right] \tag{6.12}$$

(6.11) implies

$$\sqrt{\frac{n(n+p-2)}{S^2+2c}}\,(\alpha-a) \sim t(n+p-2)$$

and Bayes estimator $\alpha^* = a$, the same as the least-square estiator of α.

Putting $p = c = 0$ implies that under the joint prior (6.9)

$$\frac{\alpha - a}{\sqrt{\frac{S^2}{n(n-2)}}} \sim t(n-2) \qquad \text{(Lindley, 1965)}$$

Similarly, integrating out (α, β) from (6.5) we obtain the marginal posterior of σ^2 given by

$$\Pi(\sigma^2 \mid D) = \frac{K_2 \exp\left[-\frac{1}{2\sigma^2}(S^2+2c)\right]}{(\sigma^2)^{(n+p-2)/2+1}} \tag{6.13}$$

$$K_2^{-1} = \int_0^\infty \frac{\exp\left[-\frac{1}{\sigma^2}\frac{(S^2+2c)}{2}\right]}{(\sigma^2)^{(n+p-2)/2+1}}\, d\sigma^2$$

$$= \frac{\Gamma\left(\frac{n+p-2}{2}\right)}{\left(\frac{S^2+2c}{2}\right)^{(n+p-2)/2}}.$$

Substituting K_2 in (6.13)

$$\Pi(\sigma^2 \mid D) = \frac{\left(\frac{S^2+2c}{2}\right)^{(n+p-2)/2} \exp\left[-\frac{S^2+2c}{2\sigma^2}\right]}{\Gamma\left(\frac{n+p-2}{2}\right)(\sigma^2)^{(n+p-2)/2+1}} \tag{6.14}$$

which implies

$\frac{S^2+2c}{\sigma^2} \sim \chi^2\ (n+p-2)$: Chi-square with $(n+p-2)$ degrees of freedom

and comparing (6.14) with the prior $h(\sigma^2)$ and following (6.3), we obtain Bayes estimator of σ^2 given by

$$(\sigma*)^2 = \frac{S^2+2c}{n+p-4}.$$

Under the joint prior (6.9) we have

$$\frac{S^2}{\sigma^2} \sim \chi^2\ (n-2) \qquad \text{(Lindley, 1965)}$$

and
$$(\sigma*)^2 = \frac{S^2}{n-4},\ n > 4.$$

6.3 Credible Interval

Equal-tail $(1-p)$-credible limits $(\theta_1,\ \theta_2)$ of a parameter θ are solutions of the equation,

$$\frac{p}{2} = \int_{-\infty}^{\theta_1} \Pi\ (\theta \mid D)\, d\theta \int_{\theta_2}^{\infty} \Pi\,(\theta \mid D)\ d\theta\ .$$

For the parameter β, we solve for $(\beta_1,\ \beta_2)$ such that

$$\frac{p}{2} = \int_{-\infty}^{\beta_1} \Pi_1\ (\beta \mid D)\ d\beta\ = \int_{\beta_2}^{\infty} \Pi_1\,(\beta \mid D)\ d\beta\ .$$

From (6.8)

$$\frac{p}{2} = \frac{1}{\sqrt{A}\ B\left(\frac{1}{2},\frac{\upsilon}{2}\right)} \int_{-\infty}^{\beta_1} \frac{d\beta}{\left[1 + \frac{(\beta-b)^2}{A}\right]^{(\upsilon+1)/2}}$$

$$= \frac{1}{\sqrt{\upsilon}\ B\left(\frac{1}{2},\frac{\upsilon}{2}\right)} \int_{-\infty}^{(\beta_1-b)\sqrt{\upsilon/A}} \frac{dt}{\left(1+\frac{t^2}{\upsilon}\right)^{(\upsilon+1)/2}}$$

$$= P\left[t < (\beta_1 - b)\sqrt{\frac{\upsilon}{A}}\right] \equiv P\left[t < -t\left(\frac{p}{2}, \upsilon\right)\right]$$

where $$(\beta_1 - b)\sqrt{\frac{\upsilon}{A}} = -t\left(\frac{p}{2}, \upsilon\right)$$

$$\beta_1 = b - \sqrt{\frac{A}{\upsilon}}\, t\left(\frac{p}{2}, \upsilon\right)$$

$$A = \frac{S^2 + 2c}{S_{xx}}, \quad \upsilon = n + p - 2.$$

Similarly,

$$(\beta_2 - b)\sqrt{\frac{\upsilon}{A}} = t\left(\frac{p}{2}, \upsilon\right),$$

$$\beta_2 = b + \sqrt{\frac{A}{\upsilon}}\, t\left(\frac{p}{2}, \upsilon\right).$$

Thus, $(1-p)$-credible limits for β are $b \pm \sqrt{\frac{A}{\upsilon}}\, t\left(\frac{p}{2}, \upsilon\right)$ which are the same as the classical confidence limits and hence, the corresponding $(1-p)$-credible interval:

$$1 - p = Pr\left\{b - \sqrt{\frac{A}{\upsilon}}\, t\left(\frac{p}{2}, \upsilon\right) \le \beta \le b + \sqrt{\frac{A}{\upsilon}}\, t\left(\frac{p}{2}, \upsilon\right)\right\}.$$

For the parameter α, we solve for (α_1, α_2) where

$$\frac{p}{2} = \int_{-\infty}^{\alpha_1} \Pi_2(\alpha \mid D)\, d\alpha = \int_{\alpha_2}^{\infty} \Pi_2(\alpha \mid D)\, d\alpha.$$

From (6.12)

$$\frac{p}{2} = \sqrt{\frac{n}{S^2 + 2c}}\; \frac{1}{B\left(\frac{1}{2}, \frac{\upsilon}{2}\right)} \int_{-\infty}^{\alpha_1} \frac{d\alpha}{\left[1 + \dfrac{(\alpha - a)^2}{\dfrac{S^2 + 2c}{n}}\right]^{\frac{(\upsilon+1)}{2}}}$$

$$= \frac{1}{\sqrt{\upsilon}\, B\left(\frac{1}{2}, \frac{\upsilon}{2}\right)} \int_{-\infty}^{L} \frac{dt}{\left(1 + \frac{t^2}{\upsilon}\right)^{(\upsilon+1)/2}}$$

$$= P\left[t < L \equiv \frac{(\alpha_1 - a)\sqrt{\upsilon}}{\sqrt{\frac{S^2 + 2c}{n}}} = -t\left(\frac{p}{2}, \upsilon\right)\right]$$

or $$\alpha_1 = a - \sqrt{\frac{S^2 + 2c}{n\upsilon}}\; t\left(\frac{p}{2}, \upsilon\right).$$

Similarly $$\alpha_2 = a + \sqrt{\frac{S^2 + 2c}{n\upsilon}}\; t\left(\frac{p}{2}, \upsilon\right).$$

Note that under the joint prior (6.9), $(1 - p)$-prediction limits of α are $a \pm \sqrt{\frac{S^2}{n(n-2)}}\; t\left(\frac{p}{2}, \upsilon\right)$, which are not the same as their classical counterparts.

Consider the credible limits on σ^2.

From (6.14) we know that

$$\frac{S^2 + 2c}{\sigma^2} \sim \chi^2(\upsilon), \quad \upsilon = n + p - 2.$$

$$\frac{p}{2} = \int_0^{\sigma^2} \Pi(\sigma^2 \mid D)\, d\sigma^2$$

$$= \frac{1}{\Gamma\left(\frac{\upsilon}{2}\right) 2^{\upsilon/2}} \int_{(S^2+2c)/\sigma^2}^{\infty} (\chi^2)^{\upsilon/2-1} \exp\left(-\frac{\chi^2}{2}\right) d\chi^2$$

$$= P\left[\chi^2 > \frac{S^2 + 2c}{\sigma^2}\right] = P\left[\frac{S^2 + 2c}{\sigma^2} > \chi_1^2\right]$$

$$= P\left[\sigma^2 < \frac{S^2 + 2c}{\chi_1^2}\right].$$

Similarly,

$$\frac{p}{2} = \int_{\sigma^2}^{\infty} \Pi\,(\sigma^2 \mid D)\; d\sigma^2$$

$$= P\left[\chi^2 < \frac{S^2 + 2c}{\sigma^2}\right]$$

$$= P\left[\frac{S^2 + 2c}{\sigma^2} < \chi_2^2\right]$$

$$= P\left[\frac{S^2 + 2c}{\chi_2^2} < \sigma^2\right].$$

Hence, the ($1 - p$)-credible interval for σ^2 is given by

$$\frac{S^2 + 2c}{\chi_2^2} < \sigma^2 < \frac{S^2 + 2c}{\chi_1^2},$$

which is the same as the classical confidence interval.

Note that $\chi^2\,(K, \upsilon) = 100\ K\%$ point of a χ^2 distribution with K degrees of freedom.

6.4 Prediction Interval

Let y be a future observation of a random variable Y for a given value of $x = x_0$.

$$y \sim N\,[\alpha + \beta\,(x_0 - \bar{x}),\ \sigma^2\,].$$

Given the data we will obtain the predictive distribution of y and construct a $(1 - p)$-prediction limits on y.

In (3.1) we have defined the predictive distribution of a future observation Y as

$$h\,(y \mid D) \propto \int_{\Omega} \Pi\,(\theta \mid D)\ f\,(y \mid \theta)\ d\theta \tag{6.15}$$

where $\Pi\ (\theta \mid D)$ is the posterior distribution of θ and $f\,(y \mid \theta)$ is the pdf of y.

For the regression model (6.1) we have obtained the joint posterior $\Pi\ (\alpha, \beta, \lambda\ \sigma^2 \mid D)$ in (6.5) under the prior (6.4). Using the independent prior (6.9)

$$h(y \mid D, x_0) \propto \iiint \frac{1}{(\sigma^2)^{(n+2)/2}} \exp\left[-\frac{1}{2\sigma^2}\{S^2 + n(\alpha - a)^2 + S_{xx}(\beta - b)^2 + (y - \alpha - \beta(x_0 - \bar{x})^2\}\right]$$

$$d\alpha\, d\beta\, d\sigma^2 \quad -\infty < \alpha < \infty, -\infty < \beta < \infty,\ \sigma^2 > 0. \qquad (6.16)$$

$$\int \exp\left[-\frac{1}{2\sigma^2}\{n(\alpha - a)^2 + (y - \alpha - \beta(x_0 - \bar{x})^2\}\right] d\alpha$$

$$= \int \exp\left[-\frac{1}{2\sigma^2}\{n(\alpha - a)^2 + (\delta - \alpha)^2\}\right] d\alpha$$

where $\delta = y - \beta(x_0 - \bar{x})$.

Completing square within the expression in the curl-brackets, we obtain

$$\exp\left[\frac{-n(\delta - a)^2}{2(n+1)\sigma^2}\right] \int \exp\left[-\frac{n+1}{2\sigma^2}\left(\alpha - \frac{an + \delta}{n+1}\right)^2\right] d\alpha$$

$$\propto \sigma \exp\left[-\frac{n}{2(n+1)\sigma^2}\{y - \beta(x_0 - \bar{x}) - a\}^2\right].$$

From (6.16)

$$h(y \mid D, x_0) \propto \frac{1}{(\sigma^2)^{(n+1)/2}} \iint \exp\left(\frac{-S^2}{2\sigma^2}\right).$$

$$\exp\left\{-\frac{S_{xx}}{2\sigma^2}(\beta - b)^2\right\}.$$

$$\exp\left[-\frac{n}{2(n+1)\sigma^2}\{y - \beta(x_0 - \bar{x}) - a\}^2\right] d\beta\, d\sigma^2. \qquad (6.17)$$

$$\exp\left[-\frac{n}{2(n+1)\sigma^2}\{y - \beta(x_0 - \bar{x}) - a\}^2\right]$$

implies that $Y \sim N\left[a + \beta(x_0 - \bar{x}),\ \sigma^2 + \frac{\sigma^2}{n}\right]$.

Integrating out α from the joint distribution of

$$Y \sim N\,[\alpha + \beta(x_0 - \bar{x}),\ \sigma^2] \text{ and } \alpha \sim N\left(a, \frac{\sigma^2}{n}\right).$$

we end up with $Y \sim N\left[\alpha + \beta\,(x_0 - \bar{x}),\, \sigma^2 + \dfrac{\sigma^2}{n}\right]$. Similarly, integrating out β from the joint distribution of $Y \sim N\left[a + \beta\,(x_0 - \bar{x}),\ \sigma^2 + \dfrac{\sigma^2}{n}\right]$ and $\beta \sim N\left[b,\ \dfrac{\sigma_2}{S_{XX}}\right]$ in (6.17) will result in $Y \sim N\left[\alpha + \beta\,(x_0 - \bar{x}),\right.$ $\left.\sigma^2 + \dfrac{\sigma^2}{n} + \dfrac{(x_0 - \bar{x})}{S_{xx}}\right] \sim N\ [a + b\,(x_0 - \bar{x}),\ L\sigma^2]$ where

$$L = 1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}.$$

From (6.17)

$$h\,(y \mid D, x_0) \propto \int \frac{1}{(\sigma^2)^{(n-1)/2+1}} \cdot$$

$$\exp\left[-\frac{1}{2\sigma^2}\left\{S^2 + \frac{(y - a - b\,(x_0 - \bar{x})^2}{L}\right\}\right] d\sigma^2 .$$

Integrating out σ^2,

$$h\,(y \mid D, x_0) \propto \left[S^2 + \frac{\{y - a - b\,(x_0 - \bar{x})\}^2}{L}\right]^{(n-1)/2}$$

$$= \frac{C}{\left[1 + \dfrac{\{y - a - b\,(x_0 - \bar{x})\}^2}{LS^2}\right]^{(\upsilon+1)/2}} . \qquad (6.18)$$

where $\upsilon = n - 2$.

Let

$$\frac{y - a - b\,(x_0 - \bar{x})}{S\,\sqrt{L}} = \frac{1}{\sqrt{\upsilon}} \qquad (6.19)$$

$$C^{-1} = S\sqrt{\frac{L}{\upsilon}} \int_{-\infty}^{\infty} \frac{dt}{\left(1 + \dfrac{t^2}{\upsilon}\right)^{(\upsilon+1)/2}}$$

$$= S \sqrt{L}\, B\left(\frac{1}{2}, \frac{\upsilon}{2}\right).$$

Substituting C in (6.18) we obtain the predictive distribution

$$h(;y \mid D, x_0) = \frac{1}{S\sqrt{L}\, B\left(\frac{1}{2}, \frac{\upsilon}{2}\right)}$$

$$\left[1 + \frac{y - a - b(x_0 - \bar{x})}{LS^2}\right]^{(\upsilon+1)/2}, \quad -\infty < y < \infty.$$

(6.19) implies

$$\sqrt{\frac{\upsilon}{L}}\left\{\frac{y - a - b(x_0 - \bar{x})}{S}\right\} \sim \text{Student's-}t \text{ with } \upsilon \text{ degrees of freedom.}$$

$(1 - p)$ prediction limits of Y are

$$a + b(x_0 - \bar{x}) \pm S\sqrt{\frac{L}{\upsilon}}\; t\left(\frac{p}{2}, \upsilon\right).$$

which is the same as the corresponding classical prediction limits.

6.5 Joint Credible region for (α, β)

We may be more interested in making a credible statement about the line itself rather than a single predicted value of $\mu_{y \mid x_0} = \alpha + \beta x_0$. This can be achieved by constructing joint credible region for (α, β). Using the prior (6.9) and integrating out σ^2 in (6.5), the joint marginal posterior of (α, β) is given by

$$\Pi_{12}(\alpha, \beta \mid D) \propto [S^2 + n(\alpha - a)^2 + S_{xx}(\beta - b)^2]^{-n/2}$$

$$\propto \left[1 + \frac{n(\alpha - a)^2 + S_{xx}(\beta - b)^2}{S^2}\right]^{-n/2}$$

$$\frac{n(\alpha - a)^2}{\sigma^2} \sim \chi^2(1), \quad \frac{S_{xx}(\beta - b)^2}{\sigma^2} \sim \chi^2(1), \quad \frac{S^2}{\sigma^2} \sim \chi^2(n-2).$$

Hence,

$$\frac{n(\alpha-a)^2+S_{xx}(\beta-b)^2}{S^2} = \frac{\dfrac{n(\alpha-a)^2+S_{xx}(\beta-b)^2}{\sigma^2}}{\dfrac{S^2}{\sigma^2}} \sim \frac{\chi^2(2)}{\chi^2(n-2)}$$

$$= \frac{2}{n-2}\left[\frac{\dfrac{\chi^2(2)}{2}}{\dfrac{\chi^2(n-2)}{n-2}}\right]$$

$$\sim \frac{\upsilon_1}{\upsilon_2} F(\upsilon_1, \upsilon_2), \quad \upsilon_1 = 2, \ \upsilon_2 = n-2.$$

Thus, $\quad \Pi(\alpha, \beta \mid D) \propto \left(1 + \dfrac{\upsilon_1}{\upsilon_2} F\right)^{(\upsilon_1+\upsilon_2)/2}, \quad 0 < F < \infty.$

The boundary of the $(1-p)$-credible band for (α, β) may be determined by solving the equation

$$n(\alpha-a)^2 + S_{xx}(\beta-b)^2 - \frac{2S^2}{n-2} F(1-p, \upsilon_1, \upsilon_2) = 0 \qquad (6.20)$$

for values of (α, β). For example,

for $\qquad \alpha = a, \ \beta = b \pm \sqrt{\dfrac{2S^2 F_0}{(n-2) S_{xx}}}$.

For $\qquad \beta = b, \ \alpha = a \pm \sqrt{\dfrac{2S^2 F_0}{n(n-2)}}$,

where $\qquad F_0 = F(1-p, \upsilon_1, \upsilon_2)$

$= 100(1-p)\%$ point of F-distribution with $\upsilon_1 = 2$, $\upsilon_2 = (n-2)$ degrees of freedom.

Plotting suitable values of (a, b) we may obtain the $(1-p)$-credible region given by the ellipse (6.20). Given the data D, to test the hypothesis H_0 : the line is given by $y = \alpha_0 + \beta_0(x - \bar{x})$ at 5% level of significance, compute $\dfrac{n-2}{2S^2}[n(\alpha_0-a)^2 + S_{xx}(\beta_0-b)^2] = F$ and reject H_0 if $F > F(.95, 2, n-2)$.

6.6 Bivariate Normal Prior

Suppose σ^2 is known and (α, β) have a bivariate normal prior. Let $E(\alpha) \equiv E(\beta) = 0$, $V(\alpha) = V_1$, $V(\beta) = V_2$ and ρ is known.

$$g(\alpha, \beta) \propto \frac{1}{\sqrt{V_1 V_2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left\{\frac{\alpha^2}{V_1} + \frac{\beta^2}{V_2} - \frac{2\rho\alpha\beta}{\sqrt{V_1 V_2}}\right\}\right]. \qquad (6.21)$$

From (6.2) and (6.22) we have the joint posterior distribution of (α, β)

$$\Pi(\alpha, \beta \mid D) = \frac{K}{\sqrt{V_1 V_2}} \exp\left[\begin{array}{l} -\dfrac{1}{2(1-p^2)}\left\{\dfrac{\alpha^2}{V_1} + \dfrac{\beta^2}{V_2} - \dfrac{2\rho\alpha\beta}{\sqrt{V_1 V_2}}\right\} + \\ \\ \dfrac{1}{2\sigma^2}\{S^2 + n(\alpha - a)^2 + S_{xx}(\beta - b)^2\} \end{array}\right]$$

$$-\infty < \alpha < \infty, \quad -\infty < \beta < \infty$$

Integrating out α and β and restoring the normalizing constants, one may obtain the marginal posteriors $\Pi_3(\alpha \mid D)$ and $\Pi_4(\beta \mid D)$ and hence, the Bayes estimators

$$\alpha^* = E(\alpha \mid D) \text{ and } \beta^* = E(\beta \mid D).$$

These estimators will require evaluation of complicated double integrals,

$$K^{-1} = \iint \Pi(\alpha, \beta \mid D)\, d\alpha\, d\beta$$

$$\alpha^* = \int \alpha \left\{\int \Pi(\alpha, \beta)\, d\beta\right\} d\alpha,$$

and

$$\beta^* = \int \beta \left\{\int \Pi(\alpha, \beta)\, d\alpha\right\} d\beta.$$

We may use instead, Lindley's approximation (5.2) to derive (α^*, β^*).

Let $u \equiv \alpha$ and $R = \log\{g(\alpha, \beta)\}$. From (5.9)

$$u(\alpha, \beta) = \left[\alpha + R_1 \sigma_{11} + R_2 \sigma_{22} + \frac{3}{2}\sigma_{11}\sigma_{12} L_{21}\right.$$

$$+ \frac{1}{2} L_{12}\left(\sigma_{11}\sigma_{22}\, 2\sigma_{11}^2\right)$$

$$\left. + \frac{1}{2}\left(\sigma_{12}\sigma_{22} L_{03} + \sigma_{11}^2 L_{30}\right)\right]_{(\hat{\alpha}, \hat{\beta})} \qquad (6.22)$$

where $(\hat{\alpha}, \hat{\beta})$ are the MLE of (α, β).

$$R_1 = \frac{\partial R}{\partial \alpha} = -\frac{1}{(1-\rho^2)}\left[\frac{\alpha}{V_1} - \frac{\beta\rho}{\sqrt{V_1 V_2}}\right]$$

$$R_2 = \frac{\partial R}{\partial \beta} = -\frac{1}{(1-\rho^2)}\left[\frac{\beta}{V_2} - \frac{\alpha\rho}{\sqrt{V_1 V_2}}\right]$$

and L_{ij}, σ_{ij} are defined in (5.2).

From (6.2)

$$L = \text{Constant} - \frac{1}{2\sigma^2}\left[S^2 + n(\alpha - a)^2 + S_{xx}(\beta - b)^2\right].$$

$$L_{11} = \frac{\partial^2 L}{\partial \alpha^2} = -\frac{n}{\sigma^2}, \quad \sigma_{11} = \frac{\sigma^2}{n},$$

$$L_{22} = \frac{\partial^2 L}{\partial \beta^2} = \frac{S_{xx}}{\sigma^2}, \quad \sigma_{22} = \frac{\sigma^2}{S_{xx}},$$

$$L_{30} = \frac{\partial^3 L}{\partial \alpha^2} = 0, \quad L_{03} = \frac{\partial^3 L}{\partial \beta^3} = 0,$$

$$L_{12} = \frac{\partial^3 L}{\partial \alpha\, \partial \beta^2} = 0, \quad L_{21} = \frac{\partial^3 L}{\partial \alpha^2\, \partial \beta} = 0$$

Substituting in (6.22)

$$\alpha * = \hat{\alpha} - \frac{1}{1-\rho^2}\left[\frac{\hat{\alpha}}{V_1} - \frac{\hat{\beta}\rho}{\sqrt{V_1 V_2}}\right].$$

Similarly,

$$\beta* = \hat{\beta} - \frac{1}{1-\rho^2}\left[\frac{\hat{\beta}}{V_2} - \frac{\hat{\alpha}\rho}{\sqrt{V_1 V_2}}\right].$$

Thus, for a suitable choice of the prior parameters (V_1, V_2, ρ) Bayes estimators $(\alpha*, \beta*)$ may be computed. (Howlader and Sinha, 1984).

6.7 Bivariate Normal Distribution

The Bivariate Normal Distribution (BVN) has received a great deal of attention (Kendall and Stuart, Vol. 2, 1973; Johnson and Kotz, 1970) although most applications of this distribution make additional assumptions about the distributional parameters, or focus on a subset of the

parameters (Lindley, 1965; Samanta and Sinha, 1979). We will apply Lindley (1980) and Tierney and Kadane (1986) — approximations to obtain Bayes estimators of the BVN parameters when all parameters are unknown.

The pdf of the BVN random variables (X_1, X_2) is given by

$$f(x_1, x_2 \mid \mu_1, \mu_2, \sigma_1, \sigma_2, \rho) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\right.$$

$$\left.\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2 - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}\right]\right\} \tag{6.23}$$

where $-\infty < x_1, x_2 < \infty$, $-\infty < \mu_1, \mu_2 < \infty$, $\sigma_1, \sigma_2 > 0$, $-1 \le \rho \le 1$.

Let $\theta = (\mu_1, \mu_2, \sigma_1, \sigma_2, \rho) \equiv (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5)$.

The MLE of the parameters are

$$\hat{\mu}_i = \bar{x}_i,\ \hat{\sigma}_i^2 = \frac{\sum_j (x_{ij}-\bar{x}_i)^2}{n}$$

$$\hat{\rho} = \frac{\sum_j (x_{1j}-\bar{x}_1)(x_{2j}-\bar{x}_2)}{n\hat{\sigma}_1\hat{\sigma}_2},\quad i = 1, 2;\ j = 1, 2, \cdots, n.$$

Assuming prior independence of the parameters $(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$, uniform priors for the means and a non-vanishing prior for ρ, Lindley (1965) suggested the joint prior

$$g(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho) \propto \frac{1}{\sigma_1\sigma_2(1-\rho^2)}. \tag{6.24}$$

Jeffreys (1961) invariant prior requires that the prior be proportional to be square-root of the determinant of Fisher's information matrix and it works out to be

$$h(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho) \propto \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)^2} \tag{6.25}$$

which is basically the square of Lindley's prior (6.24).

To avoid confusion between the population correlation coefficient ρ in (6.24) and the $\rho = \frac{\partial}{\partial\theta}\log\{g(\theta)\}$ in (5.2), let $R = \frac{\partial}{\partial\theta}\log\{g(\theta)\}$ and re-write the expansion

$$\mu^* = \left[u(\theta) + \frac{1}{2} \sum_{i=1}^{p} \sum_{j=1}^{p} \{u_{ij}(\theta) + 2u_i(\theta) R_j(\theta)\} \sigma_{ij} \right.$$

$$\left. + \frac{1}{2} \sum_{i=1}^{p} \sum_{j=1}^{p} \sum_{K=1}^{p} \sum_{l=1}^{p} L_{ijK} \, u_l \, \sigma_{ij} \, \sigma_{Kl} \right] \qquad (6.26)$$

where $\quad R_j = \dfrac{\partial}{\partial \theta_j} \{\log g(\theta)\},$

p = dimensionality of θ,

$$L = \text{constant} - n \left[\log \sigma_1 - \log \sigma_2 - \frac{1}{2} \log (1 - \rho^2) \right] - \frac{1}{2(1-\rho^2)} \cdot$$

$$\sum_{j=1}^{n} \left[\frac{(x_{1j} - \mu_1)^2}{\sigma_1^2} + \frac{(x_{2j} - \mu_2)^2}{\sigma_2^2} - \frac{2\rho}{\sigma_1 \sigma_2} . (x_{1j} - \mu_1)(x_{2j} - \mu_2) \right]$$

and u_i, u_{ij}, L_{ijk}, u_l $\hat{\theta}$ are the same as defined earlier in (5.2). We will use the prior (6.24).

After some algebra (Appendix A.8) we derive the elements of the $\left[-L_{ij}\right]_{\hat{\theta}}$ matrix given by

$$\left[-L_{ij}\right]_{\hat{\theta}} = \frac{1}{1-\rho^2} \begin{bmatrix} \dfrac{n}{\sigma_1^2} & \dfrac{-n\rho}{\sigma_1 \sigma_2} & 0 & 0 & 0 \\ \dfrac{-n\rho}{\sigma_1 \sigma_2} & \dfrac{n}{\sigma_2^2} & 0 & 0 & 0 \\ 0 & 0 & \dfrac{n(2-\rho^2)}{\sigma_1^2} & \dfrac{n(2-\rho^2)}{\sigma_2^2} & \dfrac{-n\rho}{\sigma_1} \\ 0 & 0 & \dfrac{-n\rho^2}{\sigma_1\sigma_2} & \dfrac{n(2-\rho^2)}{\sigma_2^2} & \dfrac{-n\rho}{\sigma_2} \\ 0 & 0 & \dfrac{-n\rho}{\sigma_1} & \dfrac{-n\rho}{\sigma_2} & \dfrac{n(1+\rho^2)}{(1-\rho^2)} \end{bmatrix}_{\hat{\theta}}$$

which is of the form

$$V = C \begin{bmatrix} A_{2\times2} & 0_{2\times3} \\ 0_{3\times2} & B_{3\times3} \end{bmatrix}.$$

The inverse of such a matrix

$$V^{-1} = \frac{1}{C}\begin{bmatrix} A^{-1}_{2\times 2} & 0_{2\times 3} \\ 0_{3\times 2} & B^{-1}_{3\times 3} \end{bmatrix} \qquad \text{(Anderson, 1968; pp 342).}$$

Hence,

$$[\sigma_{ij}] = [-L_{ij}]^{-1}_{\theta}$$

$$= \frac{1}{n}\begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & 0 & 0 & 0 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 & 0 & 0 & 0 \\ 0 & 0 & \dfrac{\sigma_1^2}{2} & \dfrac{\rho^2\sigma_1\sigma_2}{2} & \dfrac{\rho\sigma_1(1-\rho^2)}{2} \\ 0 & 0 & \dfrac{\rho^2\sigma_1\sigma_2}{2} & \dfrac{\sigma_2^2}{2} & \dfrac{\rho\sigma_2(1-\rho^2)}{2} \\ 0 & 0 & \dfrac{\rho\sigma_1(1-\rho^2)}{2} & \dfrac{\rho\sigma_2(1-\rho^2)}{2} & (1-\rho^2)^2 \end{bmatrix}_{\theta} \tag{6.27}$$

Inspecting the above matrix it follows that

$$\sigma_{13} = \sigma_{14} = \sigma_{15} = \sigma_{23} = \sigma_{24} = \sigma_{25} = 0. \tag{6.28}$$

Further note that $L_{ijK} = L_{iKj} = L_{jKi} = L_{jiK} = L_{Kij} = L_{Kji}$ and out of 35 distinct third derivatives L_{ijK}, only 17 different L_{ijK}'s are non-zero. The following 18 third derivatives vanish at the MLE $\hat{\theta} = (\hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}_1, \hat{\sigma}_2, \hat{\rho})$, viz.,

$$L_{111}, L_{121}, L_{122}, L_{133}, L_{135}, L_{141}, L_{143}, L_{144}, L_{155},$$
$$L_{145}, L_{222}, L_{232}, L_{233}, L_{243}, L_{244}, L_{245}, L_{253}, \text{ and } L_{255}$$

(Appendix A.9), (6.29)

$$R \propto -\log\sigma_1 - \log\sigma_2 - \log(1-\rho^2)$$

$$R_1 = \frac{\partial R}{\partial \mu_1} = 0; \quad R_2 = \frac{\partial R}{\partial \mu_2} = 0;$$

$$R_3 = \frac{\partial R}{\partial \sigma_1} = \frac{-1}{\sigma_1}; \quad R_4 = \frac{\partial R}{\partial \sigma_2} = \frac{-1}{\sigma_2}$$

and $$R_5 = \frac{\partial R}{\partial \rho} = \frac{2\rho}{1-\rho^2}.$$

For $\quad u = \mu_1,\ R_1 = 0,\ u_1 = 1,\ u_{ij} = 0.$

From (6.26)

$$\mu^*_1 = \hat{\mu} + \frac{1}{2} \sum_{i=1}^{5} \sum_{j=1}^{5} \sum_{K=1}^{5} L_{ijK}\, \sigma_{ij}\, \sigma_{Kl} = \bar{x}_1,$$

since the terms within the triple summations vanish on account of (6.28) and (6.29).

Similarly,

$$\mu^*_2 = \bar{x}_2.$$

For $\quad u = \sigma_1,\ R_1 = R_2 = 0,\ R_3 = -\dfrac{1}{\sigma_1},\ u_3 = 1,\ \mu_{3j} = 0.$

From (6.27)

$$\sum_{j=1}^{5} R_j\, \sigma_{3j} = -\frac{1}{\sigma_1}\left(\frac{\sigma_1^2}{2n}\right) - \frac{1}{\sigma_2}\left(\frac{\rho^2\sigma_1\sigma_2}{2n}\right)$$

$$+\left(\frac{2\rho}{1-\rho^2}\right)\frac{\rho\sigma_1(1-\rho^2)}{2n} = -\frac{\sigma_1(1-\rho^2)}{2n}.$$

Substituting in (6.26)

$$\sigma_1^* = \left[\sigma_1 - \frac{\sigma_1(1-\rho^2)}{2n} + \frac{1}{2}\sum_{i=1}^{5}\sum_{j=1}^{5}\sum_{K=1}^{5} L_{ijK}\,\sigma_{ij}\,\sigma_{K3}\right]_{\hat{\theta}}$$

$$= \sigma_1 - \frac{\sigma_1(1-\rho^2)}{2n} + \frac{1}{2}\,[L_{333}\,\sigma_{33}^2$$

$$+ L_{444}\,\sigma_{44}\,\sigma_{43} + L_{555}\,\sigma_{55}\sigma_{53} + L_{113}\,\sigma_{11}\,\sigma_{33} + L_{115}\,\sigma_{11}\sigma_{53}$$

$$+ L_{224}\,\sigma_{22}\,\sigma_{43} + L_{225}\,\sigma_{22}\,\sigma_{53} + L_{443}\,\sigma_{44}\,\sigma_{33} + L_{445}\,\sigma_{44}\,\sigma_{53}$$

$$+ L_{553}\,\sigma_{55}\,\sigma_{33} + L_{554}\,\sigma_{55}\,\sigma_{43}] + [L_{123}\,\sigma_{12}\,\sigma_{33} + L_{124}\,\sigma_{12}\,\sigma_{43}$$

$$+ L_{125}\,\sigma_{12}\,\sigma_{53} + L_{434}\,\sigma_{34}^2 + L_{454}\,\sigma_{45}\,\sigma_{43} + L_{535}\,\sigma_{53}^2$$

$$+ L_{545}\,\sigma_{54}\,\sigma_{53} + L_{453}\,\sigma_{45}\,\sigma_{33} + L_{345}\,\sigma_{34}\,\sigma_{53} + L_{354}\,\sigma_{35}\,\sigma_{43}]$$

$$+ \frac{3}{2}\,(L_{334}\,\sigma_{33}\,\sigma_{43} + L_{335}\,\sigma_{33}\,\sigma_{53})$$

evaluated at $\hat{\theta}$. Again all other L_{ijK} and σ_{ij} terms will vanish on account of (6.28) and (6.29).

After some algebra, we obtain

$$\sigma_1^* = \left[\sigma_1 + \frac{\sigma_1}{2n}\left(\frac{7}{2} - \rho^2\right)\right]_{\hat{\sigma}_1, \hat{\rho}}.$$

Similarly,

$$\sigma_2^* = \left[\sigma_2 - \frac{\sigma(1-\rho^2)}{2n} + \frac{1}{2}\sum_{i=1}^{5}\sum_{j=1}^{5}\sum_{K=1}^{5} L_{ijK}\,\sigma_{ij}\,\sigma_{K4}\right]_{\theta}$$

$$= \left[\sigma_2 - \frac{\sigma_2(1-\rho^2)}{2n} + \left\{\frac{1}{2}\left(L_{113}\,\sigma_{11}\,\sigma_{34} + L_{115}\,\sigma_{11}\,\sigma_{54}\right.\right.\right.$$

$$+ L_{224}\,\sigma_{22}\,\sigma_{44} + L_{225}\,\sigma_{24}\,\sigma_{54} + L_{335}\,\sigma_{33}\,\sigma_{34}$$

$$+ L_{444}\,\sigma_{44}^2 + L_{555}\,\sigma_{55}\,\sigma_{54}$$

$$+ L_{334}\,\sigma_{33}\,\sigma_{44} + L_{335}\,\sigma_{33}\,\sigma_{54} + L_{443}\,\sigma_{44}\,\sigma_{34}$$

$$\left.\left.+ L_{445}\,\sigma_{44}\,\sigma_{54} + L_{553}\,\sigma_{55}\,\sigma_{34} + L_{554}\,\sigma_{55}\,\sigma_{44}\right)\right\}$$

$$+ L_{123}\,\sigma_{12}\,\sigma_{34} + L_{125}\,\sigma_{12}\,\sigma_{54} + L_{124}\,\sigma_{12}\,\sigma_{44} + L_{343}\sigma_{34}^2$$

$$+ L_{353}\,\sigma_{35}\,\sigma_{34} + L_{434}\,\sigma_{43}\,\sigma_{44} + L_{454}\,\sigma_{45}\,\sigma_{44}$$

$$\left[+ L_{535}\,\sigma_{53}\,\sigma_{54} + L_{545}\,\sigma_{54}^2 + 2L_{345}\,\sigma_{34}\,\sigma_{54}\right]_{\theta}$$

$$= \left[\sigma_2 + \frac{\sigma_2}{2n}\left(\frac{7}{2} - \rho^2\right)\right]_{\hat{\theta}_2, \hat{\rho}}.$$

$$\rho^* = \left[\rho + \frac{\rho(1-\rho^2)}{2n} + \frac{1}{2}\sum_{i=1}^{5}\sum_{j=1}^{5}\sum_{K=1}^{5} L_{ijK}\,\sigma_{ij}\,\sigma_{K5}\right]_{\theta}$$

$$= \left[\rho + \frac{\rho(1-\rho^2)}{2n} + \left\{\frac{1}{2}\left(L_{113}\,\sigma_{11}\,\sigma_{35} + L_{115}\,\sigma_{11}\,\sigma_{55}\right.\right.\right.$$

$$+ L_{224}\,\sigma_{22}\,\sigma_{45} + L_{225}\,\sigma_{22}\,\sigma_{55} + L_{333}\,\sigma_{33}\,\sigma_{35} + L_{444}\,\sigma_{44}\,\sigma_{45}$$

$$+ L_{555}\,\sigma_{55}^2 + L_{334}\,\sigma_{33}\,\sigma_{45} + L_{335}\,\sigma_{33}\,\sigma_{55} + L_{443}\,\sigma_{44}\,\sigma_{35}$$

$$\left.\left.+ L_{445}\,\sigma_{44}\,\sigma_{55} + L_{553}\,\sigma_{55}\,\sigma_{35} + L_{554}\,\sigma_{55}\,\sigma_{45}\right)\right\}$$

$$+ L_{123}\,\sigma_{12}\,\sigma_{35} + L_{124}\,\sigma_{12}\,\sigma_{45}$$

$$+ L_{125}\sigma_{12}\sigma_{55} + L_{343}\sigma_{34}\sigma_{35} + L_{353}\sigma_{35}^2$$

$$+ L_{434}\sigma_{43}\sigma_{45} + L_{454}\sigma_{45}^2 + L_{535}\sigma_{53}\sigma_{55}$$

$$+ L_{545}\sigma_{54}\sigma_{55} + L_{354}\sigma_{35}\sigma_{45} + L_{345}\sigma_{34}\sigma_{55} + L_{543}\sigma_{54}\sigma_{35}\Big]_{\theta}$$

$$= \left[\rho - \frac{3\rho(1-\rho^2)}{2n}\right]_{\hat{\rho}}.$$

We now consider Tierney-Kadane (T–K) approximation (1986) given by (5.34):

$$u^* = \left[\frac{\det \Sigma_*}{\det \Sigma_0}\right]^{1/2} \exp\left[n\{L_*(\theta_*) - L_0(\theta_0)\}\right]$$

where, $\quad L_0(\theta) = \dfrac{1}{n}[L(\theta) + \log\{g(\theta)\}]$

$$L_*(\theta) = L_0(\theta) + \frac{1}{n}\log u(\theta),$$

θ_0 = maximizing point of L_0,

θ_* = maximizing point of L_*

$$\Sigma_0 = -\left[L_{ij}\right]_{\theta_0}^{-1}, \quad \Sigma_* = -\left[L_{ij}\right]_{\theta_*}^{-1}$$

we have,

$$L_0(\theta) = \text{Constant} + \frac{1}{n}\log\left[\frac{1}{\sigma_1\sigma_2(1-\rho^2)}\right]$$

$$-\frac{1}{2}\left[\log\sigma_1^2 + \log\sigma_2^2 + \log(1-\rho^2)\right]$$

$$-\frac{1}{2n(1-\rho^2)}\cdot$$

$$\sum_{i=1}^{n}\left[\frac{(x_{1i}-\mu_1)^2}{\sigma_1^2} + \frac{(x_{2i}-\mu_2)^2}{\sigma_2^2} - \frac{2\rho(x_{1i}-\mu_1)(x_{2i}-\mu_2)}{\sigma_1\sigma_2}\right]$$

$$L_*(\theta) = L_0(\theta) + \frac{1}{n}\log u(\theta).$$

As we have already remarked, the maxima for $L_{(i)}$ will be the same regardless of the parameter being estimated, but L_* has to be recalculated and a new maximum has to be obtained for each estimating parameter. When these maximum values are substituted in the matrices $\Sigma_{(i)}$ and Σ_* a closed form solution becomes almost intractable and numerical computation routines have to be used. Newton's method is usually sufficient to obtain the maxima, using the MLE as the starting point.

Numerical Example

Bivariate measures on reading skills were taken on 26 children dealing with the Woodcock reading test and a bivariate normal model was used (Pope, Lehrer and Stevens, 1980). The following table presents the estimated parameters. The entries within the parenthesis represents the estimated variances.

Parameters	*MLE*	*LINDLEY*	*T–K*
μ_1	6.4038	6.4038	6.3775
	(0.1292)	(0.1631)	(0.1208)
μ_2	6.8692	6.8692	6.7971
	(0.2897)	(0.2673)	(0.2656)
σ_1	2.0591	2.1790	2.1911
	(0.0702)	(0.0672)	(0.0577)
σ_2	2.6362	2.7896	2.7644
	(0.1586)	(0.1101)	(0.0938)
ρ	0.6877	0.6668	0.6702
	(0.0234)	(0.0102)	(–0.0288)

Both approximations produce similar results. The estimated variances are uniformly in the order MLE > Lindley > T–K except for μ_1. The variance estimator for the estimated r turns out to be negative for the T–K method which is not an unusual phenomenon for small samples (Tierney & Kadane, 1986), (Sloan & Sinha, 1990, 1993).

Exercises

1. Show that for the BVN model (6.23) Jeffreys (1961) invariant prior is given by

$$g(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho) \propto \frac{1}{\sigma_1^2 \sigma_2^2 (1-\rho^2)^2}$$

[The following problems refer to Lindley's prior (6.24) and Lindley's approximation (5.2)]

2. A random sample of size $n = 40$ was drawn from a BVN $(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$. Obtain Bayes estimators of the parameters.

Data :

X_1	X_2
24.9588	33.0542
20.5262	26.4479
18.2001	25.6546
10.9389	11.3420
15.9053	25.0084
25.4497	36.5334
18.6799	24.2682
18.8301	25.1105
21.9978	28.4458
16.7501	27.0237
22.8629	27.2610
23.1557	24.2802
21.9560	24.7059
20.7036	31.6767
22.8441	30.6286
22.2317	24.7393
18.2507	23.2752
13.7272	8.2495
23.6171	30.7795
23.8984	30.8680

Data:

X_1	X_2
21.7357	35.0754
12.2660	16.8193
19.0728	27.7894
18.6102	32.3981
19.6238	30.3515
14.3285	15.2181
17.1007	20.5092
20.5917	26.1282
21.3338	23.9216
17.6428	22.9243
20.7116	21.9542
23.7759	33.0239
18.5458	20.6034
23.0135	33.6604
23.3662	32.5024
22.1166	26.0987
23.6663	26.8655
22.9079	32.3460
17.4686	19.6732
20.4957	24.5349

3. Suppose the following data represent a random sample of size $n = 40$ from a BVN $(\mu_1, \mu_2, \sigma^2, \sigma^2, \rho)$. The MLE of the parameters are given by

$$\hat{\mu}_1 = \bar{x}_1, \hat{\mu}_2 = \bar{x}_2, \hat{\sigma}^2 = \frac{\sum_{i=1}^{n} (x_{1i} - \bar{x}_1)^2 + \sum_{i=1}^{n} x_{2i} - \bar{x}_2)^2}{2n}$$

and, $$\hat{\rho} = \frac{\sum_{i=1}^{n} (x_{1i} - \bar{x})(x_{2i} - \bar{x}_2)}{2\hat{\sigma}^2} \quad \text{(Appendix A.10)}$$

Compute Bayes estimator of $(\mu_1, \mu_2, \sigma^2, \rho)$.

Data:

X_1	X_2
21.0307	27.7719
16.5292	22.1405
20.1845	26.4351
14.3735	21.4112
23.9516	25.4843
25.2371	29.1367
23.3485	28.2456
17.5661	26.9202
19.0131	29.1927
19.0623	24.0620
20.2985	26.8727
19.9754	26.1199
21.3133	27.5033
21.5763	30.0295
17.2177	23.4353
16.5213	21.4554
19.9660	26.4064
22.0274	25.4624
20.1692	25.8011
18.8944	23.6987

Data:

X_1	X_2
18.6354	26.3672
21.6339	27.5634
20.8216	24.3468
18.3045	24.8814
20.9327	33.0667
21.4003	23.5608
22.7860	24.2742
19.9146	27.7417
17.6380	25.9551
17.2995	26.4940
20.8529	25.6493
24.0089	26.4716
23.6006	29.0022
13.0345	18.7012
20.4407	21.2737
21.9393	31.9826
22.4331	29.8905
16.6187	23.2174
21.3873	22.9700
22.3191	28.0676

4. A random sample of size $n = 40$ was drawn from a BVN $(\mu, \mu, \sigma^2, \sigma^2, \rho)$. The MLE of the parameters are given by

$$\hat{\mu} = \frac{\bar{x}_1 + \bar{x}_2}{2} \quad \hat{\sigma}^2 = \frac{\sum_{i=1}^{n} (x_{1i} - \hat{\mu})^2 + \sum_{i=1}^{n} (x_{2i} - \hat{\mu})^2}{2n}$$

and $$\hat{\rho} = \frac{\sum_{i=1}^{n} (x_{1i} - \hat{\mu})(x_{2i} - \hat{\mu})}{n\hat{\sigma}^2} \quad \text{(Appendix A.11).}$$

Compute Bayes estimators of (μ, σ^2, ρ).

Data:

X_1	X_2
20.6910	20.0935
15.4368	18.3902
23.8797	22.6714
15.9042	20.1698
19.5060	18.4900
18.1903	17.5238
24.4615	27.8024
16.3960	18.3274
20.2832	19.7576
23.7312	16.5250
17.5054	15.8553
24.7685	21.4703
24.6750	24.0349
14.5577	16.2587
20.4414	20.4763
19.2262	24.2486
20.0187	18.7531
16.9974	23.6831
21.1060	21.6315
24.2289	24.8988

Data:

X_1	X_2
20.4243	21.3455
20.4323	19.7038
17.0910	20.6181
16.1771	15.3523
18.4237	14.5709
24.8420	25.0089
19.8321	19.3183
15.7629	17.2150
19.4805	17.3882
19.0189	17.5752
16.3926	23.3315
24.3705	20.1044
21.6462	18.4886
26.1048	21.9950
19.0563	19.9932
18.5065	18.0804
19.9094	15.4539
23.2575	22.4698
16.7991	16.9647
19.6867	20.4464

CHAPTER

7 Mixture Distributions*

In life testing problems one often encounters situations where the underlying distributions are not homogeneous. They may consist of several subpopulations mixed in unknown proportions. A warehouse of electrical components may be comprised of two subpopulations, each with a different life expectancy. A manufacturer may produce a brand name line as well as a generic product. Mendenhall and Hader (1958) observed failure times of transmitter-receivers until a specified time T had elapsed. When brought in for maintenance, the items were further checked and classified into two groups — confirmed failures and unconfirmed failures. Thus, once a failure occurred, each failed item may be attributed to the appropriate subpopulation.

In this chapter we will obtained Bayes estimators of the parameters of the mixture of two exponentials and Weibull distributions.

7.1 Mixture of Exponential Distributions

We consider the mixture distribution,

$$F(t) = p\,F_1(t) + qF_2(t), \quad 0 < p < 1, \; q = 1 - p$$

where $$F_i(t) = 1 - \exp\left(-\frac{t}{\theta_i}\right), \quad i = 1, 2$$

*May be omitted on first Reading.

with the corresponding density function

$$f(t \mid \theta_1, \theta_2, p) = \frac{p}{\theta_1} \exp\left(-\frac{t}{\theta_1}\right) + \frac{q}{\theta_2} \exp\left(-\frac{t}{\theta_2}\right). \tag{7.1}$$

Suppose n units from the Mendenhall and Hader (1958) mixture model (7.1) are subjected to some life testing experiment and the experiment is terminated after a pre-assigned time T hours have elapsed. Suppose r units have failed during the interval $(0, T)$, r_1 units from the first subpopulation, r_2 units from the second subpopulation, $r = r_1 + r_2$ and $(n - r)$ units are still functioning.

Let t_{ij} denote the failure time of the j^{th} unit belonging to the i^{th} subpopulation, $t_{ij} \le T$, $j = 1, 2, \ldots, r_i$, $i = 1, 2$. The likelihood function,

$$L\ (p, \theta_1, \theta_2 \mid t)$$

$$\propto p^{r_1} q^{r_2} \prod_{j=1}^{r_1} \left\{\frac{1}{\theta_1} \exp\left(-\frac{t_{1j}}{\theta_1}\right)\right\}.$$

$$\prod_{j=1}^{r_2} \left\{\frac{1}{\theta_2} \exp\left(-\frac{t_{2j}}{\theta_2}\right)\right\} \cdot \left[\int_T^{\infty} f(t \mid p, \theta_1, \theta)\, dt\right]^{n-r}$$

where $t = (t_{11}\ t_{12}, \ldots, t_{1r_1}, t_{21}, t_{22}, \ldots, t_{2r_2})$.

Let

$x_{ij} = \frac{t_{ij}}{T}$, $\beta_i = \frac{\theta_i}{T}$ and $\{x_{11}, x_{12}, \cdots, x_{1r_1}; x_{21}, x_{22}, \cdots, x_{2r_2}\} = \underline{x}$ be the data.

We may write the likelihood function

$$L\,(\beta_1, \beta_2 \mid \underline{x}) \propto \left(\frac{p}{\beta_1}\right)^{r_1} \left(\frac{q}{\beta_2}\right)^{r_2} \left\{\exp\left(\frac{-r_1 \bar{x}_1}{\beta_1} - \frac{r_2 \bar{x}_2}{\beta_2}\right)\right\}.$$

$$\left\{p \,.\exp\left(-\frac{1}{\beta_1}\right) + q.exp\left(-\frac{1}{\beta_2}\right)\right\}^{(n-r)} \tag{7.2}$$

where $r_1 \bar{x}_1 = \sum_{j=1}^{r_1} x_{1j}$, $r_2 \bar{x}_2 = \sum_{j=1}^{r_2} x_{2j}$.

Note that for the exponential model (7.1), we have

$$f_i\,(x \mid \beta_i) = \frac{1}{\beta_i} \exp\left(\frac{-x}{\beta_i}\right)$$

and Jeffreys' invariant prior $g_i(\beta_i) \propto \dfrac{1}{\beta_i}$.

Assuming (β_i, p) are independent *a-priori* and $U(0, 1)$ *prior* for p, the joint prior distribution

$$g(\beta_1, \beta_2, p) \propto \frac{1}{\beta_1 \beta_2}. \tag{7.3}$$

From (7.2) and (7.3) we have the joint posterior distribution,

$$\Pi(\beta_1, \beta_2, p \mid \underline{x}) \propto \frac{p^{r_1}(1-p)^{r_2}}{\beta_1^{r_1+1}\ \beta_2^{r_2+1}} \left\{ \exp\left(\frac{-r_1 \bar{x}_1}{\beta_1} - \frac{r_2 \bar{x}_2}{\beta_2} \right) \right\} \sum_{k=0}^{n-r} \binom{n-r}{k}.$$

$$\left\{ (1-p) \exp\left(-\frac{1}{\beta_2} \right) \right\}^k \left\{ p \exp\left(-\frac{1}{\beta_1} \right) \right\}^{n-r-k}$$

$$\propto \sum_{k=0}^{n-r} \binom{n-r}{k} p^{n-k-r_2} (1-p)^{r_2+k} \exp\left\{ - \frac{\left(\dfrac{n-k-r+r_1\bar{x}_1}{\beta_1} \right)}{\beta_1^{r_1+1}} \right\}.$$

$$\left\{ \exp \frac{\left(-\dfrac{r_2 \bar{x}_2 + k}{\beta_2} \right)}{\beta_1^{r_2+1}} \right\}.$$

Integrating out β_2, p and β_1 in turn we obtain the marginal posterior distribution of β_1 given by

$$\Pi(\beta_1 \mid \underline{x}) = C \sum_{k=0}^{n-r} \binom{n-r}{k} \frac{B(n-k-r_2+1, r_2+k+1)}{(r_2 \bar{x}_2 + k)^{r_2}}.$$

$$\left[\frac{\exp\left(-\dfrac{n-r-k+r_1\bar{x}}{\beta_1} \right)}{\beta_1^{r_1+1}} \right] \tag{7.4}$$

where, $\quad C^{-1} = \Gamma(r_1) \displaystyle\sum_{k=0}^{n-r} \binom{n-r}{k} \frac{B(n-k-r_2+1, r_2+k+1)}{(r_2\bar{x}_2+k)^{r_2} (n-r-k+r_1\bar{x}_1)^{r_1}}.$

Similarly,

$$\Pi(\beta_2 \mid \underline{x}) = \frac{C\Gamma(r_1)}{\Gamma(r_2)} \sum_{k=0}^{n-r} \binom{n-r}{k} \cdot \frac{B(n-k-r_2+1,\, r_2+k+1)}{(n-r-k+r_1\bar{x}_1)^{r_1}} \left[\frac{\exp\left(-\dfrac{r_2\bar{x}_2+k}{r_2}\right)}{r_2^{r_2+1}} \binom{n-r}{k} \right] \tag{7.5}$$

and,

$$\Pi(p \mid \underline{x}) = C\Gamma(r_1) \sum_{k=0}^{n-r} \frac{p^{n-k-r_2}(1-p)^{r_2+k}}{(n-k-r+r_1\bar{x}_1)^{r_1}\,(r_2\bar{x}_2+k)^{r_2}} \cdot \binom{n-r}{k}$$

Corresponding Bayes' estimates of the parameters are,

$$r_1^* = C\Gamma(r_1-1) \sum_{k=0}^{n-r} \binom{n-r}{k} \frac{B(n-k-r_2+1,\ r_2+k+1)}{(r_2\bar{x}_2+k)^{r_2}\ (n-r-k+r_1\bar{x}_1)^{r_1-1}}, \tag{7.6}$$

$$r_2^* = \frac{C\Gamma(r_1)}{r_2-1} \sum_{k=0}^{n-r} \binom{n-r}{k} \frac{B(n-k-r_2+1,\ r_2+k+1)}{(r_2\bar{x}_2+k)^{r_2-1}\ (n-r-k+r_1\bar{x}_1)^{r_1}}, \tag{7.7}$$

$$p* = C\Gamma(r_1) \sum_{k=0}^{n-r} \binom{n-r}{k} \frac{B(n-k-r_2+2,\ r_2+k+1)}{(n-k-r+k_1\bar{x}_1)^{r_1}\ (r_2\bar{x}_2+k)^{r_2}}.$$

$$E(r_1^2+\underline{x}) = C\Gamma(r_1-2) \sum_{k=0}^{n-r} \frac{B(n-k-r_2+1,\, r_2+k+1)}{(r_2\bar{x}_2+k)^{r_2}\,(n-r-k+r_1\bar{x}_1)^{r_1-2}}. \tag{7.8}$$

We may now obtain the posterior variance,

$$V(r_1 \mid \underline{x}) = E(r_1^2 \mid \underline{x}) - \{E(r_1 \mid \underline{x})\}^2$$

from (7.6) and (7.8), and similarly for $V(\beta_2 \mid \underline{x})$ and $V(p \mid \underline{x})$.

Example 7.1:

We refer to the life testing data due to Mendenhall and Hader (1958). The test was terminated at $T = 630$ hours, $n = 369$, $r_1 = 107$, $r_2 = 218$, $\bar{x}_1 = 0.3034862$, $\bar{x}_2 = 0.3644677$. Mendenhall and Hader (1958) used numerical iteration to obtain the Maximum Likelihood Estimates (MLE) of the parameters (θ_1, θ_2, p) and computed the large-sample variance-covariance of these estimates. Bayes estimato:s (7.6) – (7.8) are given in closed forms and are easy to compute.

Table 7.1 — MLE and Bayes estimates using Mendenhall and Hader data

Parameters	*MLE*	*Estimated Variance*	*Bayes Estimates*	*Posterior Variance*
p	0.3098	0.0007	0.3130	0.0007
β_1	0.3718	0.0028	0.3829	0.0029
β_2	0.5328	0.0017	0.5315	0.0017

With large n, r_1, r_2 such as we have, the estimates are very close which is expected. (Sinha 1983).

For a complete sample case, put $r = n$ in (7.6) – (7.8). We obtain

$$\theta_1^* = \frac{\sum_{j=1}^{n_1} t_{1j}}{n_1 - 1}, \quad \theta_2^* = \frac{\sum_{j=1}^{n_2} t_{2j}}{n_2 - 1} \quad \text{and} \quad p^* = \frac{n_1 + 1}{n + 2} \tag{7.9}$$

where $\quad t_{ij} = T x_{ij}$, $\beta_1 = T\theta_i$ and $n = n_1 + n_2$.

7.2 Bayesian Prediction Interval

Re-writing (7.2), the likelihood function,

$$L(\theta_1, \theta_2, p \mid \underline{t}) \propto \sum_{k=0}^{n-r} \binom{n-r}{k} p^{n-k-r_2} (1-p)^{r_2+k} \cdot$$

$$\exp\left\{-\frac{r_1 \bar{t}_1 + (n-r-k)\,T}{\theta_1}\right\} \cdot$$

$$\exp\left\{-\frac{r_2 \bar{t}_2 + kT}{\theta_2}\right\} \theta_1^{-(r_1+1)}\, \theta_2^{-(r_2+1)} \tag{7.10}$$

where $\quad r_1\bar{t}_1 = \sum_{j=1}^{r_1} t_{1j}$, $r_2\bar{t}_2 = \sum_{j=1}^{r_2} t_{2j}$.

Consider a $U(0, 1)$ prior for p and an inverted gamma prior for θ_1 with parameters (a_i, b_i), $i = 1, 2$.

Assuming that the parameters (p, θ_1, θ_2) are *a-priori* independent, the joint prior

$$g(\theta_1, \theta_2, p) \propto \exp\left\{-\left(\frac{a_1}{\theta_1} + \frac{a_2}{\theta_2}\right)\right\} \theta_1^{-(b_1+1)}\, \theta_2^{-(b_2+1)},$$

$$a_1,\ a_2,\ b_1,\ b_2,\ \theta_1,\ \theta_2 > 0\ ,\ 0<p<1\ . \tag{7.11}$$

Combining (7.10) and (7.11), we have the joint posterior distribution,

$$\Pi(\theta_1, \theta_2, p \mid \underline{t}) \propto \sum_{k=0}^{n-r} \binom{n-r}{k} \left[\exp\left\{-\frac{r_1 \bar{t}_1 + (n-r-k)\,T + a_1}{\theta_1}\right\} \cdot \exp\left\{-\frac{r_2 \bar{t}_2 + kT + a_2}{\theta_2}\right\} \theta_1^{-(r_1+b_1+1)}\ \theta_2^{-(r_2+b_2+1)} \right].$$

The predictive distribution of a future observation Y is given by

$$h(y \mid \underline{t}) \propto \int_0^\infty \int_0^\infty \int_0^1 \Pi(\theta_1, \theta_2, p \mid \underline{t})\ f(y \mid p, \theta_1, \theta_2)\ dp\ d\theta_1\ d\theta_2$$

$$\propto \sum_{k=0}^{n-r} \binom{n-r}{k} \left[\frac{\Gamma(r_1+b_1+1)\,\Gamma(r_2+b_2)}{\{r_1\bar{t}_1 + (n-r-k)T + a_1 + y\}^{r_1+b_1+1}\ \{r_2\bar{t}_2 + kT + a_2\}^{r_2+b_2}} \cdot \int_0^1 p^{n-k-r_2+1}\,(1-p)^{r_2-k}\,dp \right.$$

$$\left. + \frac{\Gamma(r_1+b_1)\ \Gamma(r_2+b_2+1)}{\{r_1\bar{t}_1 + (n-r-k)\,T + a_1\}^{r_1+b_1}\ \{r_2\bar{t}_2 + kT + a_2 + y\}^{r_2+b_2+1}} \cdot \int_0^1 p^{n-k-r_2} \cdot (1-p)^{r_2+k+1}\,dp \right]$$

and we have

$$h(y \mid t) = C \sum_{k=0}^{n-r} \binom{n-r}{k} \cdot$$

$$\left[\frac{(r_1+b_1)\,B(n-k-r_2+2,\ r_2+k+1)}{\{r_1\bar{t}_1 + (n-r-k)\,T + a_1 + y\}^{r_1+b_1+1}\ \{r_2\bar{t}_2 + kT + a_2\}^{r_2-b_2}} + \frac{(r_2+b_2)\,B(n-k-r_2+1,\ r_2+k+2)}{\{r_1\bar{t}_1 + (n-r-k)\,T + a_1\}^{r_1+b_1}\ \{r_2\bar{t}_2 + kT + a_2 + y\}^{r_2+b_2+1}} \right] \tag{7.12}$$

where

$$C^{-1}=\sum_{k=1}^{n-r}\binom{n-r}{k}\frac{B(n-k-r_2+1,\ r_2+k+1)}{\{r_1\bar{t}_1+(n-r-k)T+a_1\}^{r_1+b_1}\ \{r_2\bar{t}_2+kT+a_2\}^{r_2+b_2}}.$$

The 100 (1 − α)% predictive limits (*L*, *U*) are solutions of the equations,

$$\int_0^L h(y\mid t)\,dy=\frac{\alpha}{2}=\int_U^\infty h(y\mid t)\,dy.$$

From (7.12) we obtain equations for the predictive limit *L* and *U* given by

$$\frac{\alpha}{2}=C\sum_{k=0}^{n-r}\binom{n-r}{k}\Bigg[\frac{B(n-r_2-k+2,\ r_2+k+1)}{\{r_2\bar{t}_2+kT+a_2\}^{r_2+b_2}}.$$

$$\left\{\frac{1}{\{r_1\bar{t}_1+(n-r-k)T+a_1\}^{r_1+b_1}}-\frac{1}{\{r_1\bar{t}_1+(n-r-k)T+a_1+L\}^{r_1+b_1}}\right\}$$

$$+\frac{B(n-r_2-k+1,\ r_2+k+2)}{\{r_1\bar{t}_1+(n-r-k)T+a_1\}^{r_1+b_1}}.$$

$$\left\{\frac{1}{\{r_2\bar{t}_2+kT+a_2\}^{r_2+b_2}}-\frac{1}{\{r_2\bar{t}_2+kT+a_2+L\}^{r_2+b_2}}\right\}\Bigg]$$

and,

$$\frac{\alpha}{2}=C\sum_{k=0}^{n-r}\binom{n-r}{k}.$$

$$\left[\frac{B(n-r_2-k+2,\ r_2+k+1)}{\{r_2\bar{t}_2+kT+a_2\}^{r_2+b_2}\ \{r_1\bar{t}_1+(n-r-k)T+a_1+U\}^{r_1+b_1}}\right.$$

$$\left.+\frac{B(n-r_2-k+1,\ r_2+k+2)}{\{r_1\bar{t}_1+(n-r-k)T+a_1\}^{r_1+b_1}\ \{r_2\bar{t}_2+kT+a_2+U\}^{r_2+b_2}}\right] \qquad (7.13)$$

which can be solved by an iterative linear search technique.

Example 7.2:

Suppose the following data represent a random sample of size $n = 100$ from the mixture distribution (7.1) truncated at $T = 100$.

Subpopulation 1:

13.8988	13.6771	51.5283	66.2278	40.3650	8.6456	103.4692
0.2087	45.5953	9.9463	24.5634	48.6234	15.6106	35.5825
48.9956	9.6344	48.2052	5.1446	27.7656	200.8856	27.3509
45.1505	124.9026	43.1305	24.0437	11.4011	3.8054	131.3938
34.4975	10.6014	23.5108	65.1220	80.3535	3.0100	31.5811
18.4381	90.6015	49.0108	15.2257	99.5481	4.7702	50.7063
34.7668	104.8867	157.7569	14.9665	71.6285	10.4170	43.6357
2.2476	88.6533	41.4672	82.7512	79.6639	21.5347	11.6312
99.4277	111.6388	64.7181	2.8014	6.4945	56.7423	33.7771
60.7091	64.2133					

Subpopulation 2;.

20.7096	99.8087	45.8027	9.8603	74.2290	30.6392	13.6507
7.7776	93.3770	5.1013	23.2716	2.5084	15.0578	6.9616
44.9577	47.4253	3.0741	12.8670	24.2991	20.2314	184.8992
16.0200	13.2814	2.5325	67.7229	1.1076	8.5718	5.9654
15.7602	21.8975	33.9987	67.8506	4.2815	81.6555	29.9614

We observe $r_1 = 58,\ r_2 = 34,\ \bar{t}_1 = 36.905,\ \bar{t}_2 = 28.595$.

From (7.13) we obtain the 95% Predictive Intervals δ for various values of prior parameters $(a_i,\ b_i)$.

		(b_1, b_2)				
	(L,U)	(10,10)	(20,20)	(30,30)	(40,40)	(50,50)
	(10,10)	(.67,125.62)	(.46,94.37)	(.35,75.60)	(.28,63.05)	(.24,54.08)
	δ	124.95	93.91	75.25	62.77	53.84
	(20,,20)	(.69,126.82)	(.48,95.25)	(.36,76.30)	(.29,63.64)	(.25,54.58)
	δ	126.13	94.77	75.94	63.35	54.33
(a_1,a_2)	(30,30)	(.71,128.03)	(.40,96.13)	(.37,77.00)	(.30,64.23)	(.25,55.09)
	δ	127.32	95.64	76.63	63.93	54.84

(40,40)	(.73,129.25)	(.50,97.01)	(.38,77.71)	(.31,64.81)	(.26,55.59)
δ	128.52	96.51	77.33	64.50	55.33
(50,50)	(.74,130.48)	(.51,97.89)	(.39,78.41)	(.32,65.40)	(.27,56.09)
δ	129.74	97.38	78.02	65.08	55.82

Shifts in the b_i parameters cause much more movement in the predictive interval than shifts in the a_1 parameters. Because the predictive distribution is unimodal and becomes steeper with larger b_i values, the length of the predictive interval decreases. Furthermore, the upper limit of the interval (U) moves more readily than the lower limit (L) because of the shape of the distribution.

As expected, setting the prior parameters to zero is equivalent to a state of ignorance of the parameters. This produces extremely wide predictive intervals (1.04, 168.24) of width $\delta = 167.20$.

(Sloan and Sinha, 1990).

7.3 Mixture of Weibull distributions

We consider the mixture of Weibull distributions,

$$F(t) = \alpha F_1(t) + (1-\alpha) F_2(t) \tag{7.14}$$

where,

$$F_i(t) = 1 - \exp\left(-\frac{t^{p_i}}{\theta_i}\right), \quad i = 1, 2.$$

Under the same setup as in the mixture of exponential distributions, the likelihood function is given by:

$$L(p_1, p_2, \theta_1, \theta_2, \alpha \mid \underline{t}) \propto \frac{\alpha^{r_1}(1-\alpha)^{r_2} p_1^{r_1} p_2^{r_2} \lambda_1^{p_1} \lambda_2^{p_2}}{\theta_1^{r_1} \theta_2^{r_1}} \cdot$$

$$\left[\exp\left\{-\left(\sum_{j=1}^{r_1} t_{1j}^{p_1}/\theta_1 + \sum_{j=1}^{r_2} t_{2j}^{p_2}/\theta_2\right)\right\}\right] \cdot$$

$$\left[\alpha \exp\left(-T^{p_1}/\theta_1\right) + (1-\alpha)\exp(-T^{p_2}/\theta_2\right]^{(n-r)},$$

$$0 < \alpha < 1; \quad \theta_1, \theta_2, p_1, p_2 > 0$$

and $\quad \lambda_1 = \prod_{j=1}^{r_1} t_{1j}, \ \lambda_2 = \prod_{j=1}^{r_2} t_{2j}.$

Let us assume that int from 0 to u *a-priori* $(\theta_1, \theta_2, p_1, p_2, \alpha)$ are independent, $\alpha \sim \cup(0, 1)$, and we are in a state of in-ignorance about p_i, θ_1 and θ_2 so that Jeffreys (1961) 'vague' prior would be appropriate. Hence, the joint prior

$$g(\theta_1, \theta_2, p_1, p_2, \alpha) \propto \frac{1}{\theta_1 \theta_2}. \tag{7.15}$$

The joint posterior distribution,

$$\Pi(\theta_1, \theta_2, p_1, p_2, \alpha \mid \underline{t}) \propto p_1^{r_1} p_2^{r_2} \lambda_1^{p_1} \lambda_2^{p_2} \sum_{k=0}^{n-r} \binom{n-r}{k} \alpha^{n-k-r_2} (1-\alpha)^{r_1+k}.$$

$$\left[\exp\left\{ \frac{\sum_{j=1}^{r_1} t_{1j}^{p_1} + (n-r-k)^{p_1} T^{p_1}}{\theta_1} \right\} \Big/ \theta_1^{r_1+1} \exp\left\{ \frac{\sum_{j=1}^{r_2} t_{2j}^{p_2} + kT^{p_2}}{\theta_2} \right\} \Big/ \theta_2^{r_2+1} \right] \tag{7.16}$$

Let $\quad x_{ij} = \frac{t_{ij}}{T}, \quad \underline{x} = (x_{11}, x_{12}, \ldots, x_{1r_1}; \ \ x_{21}, x_{22}, \ldots, x_{2r_2})$

and $\quad \delta_1 = \prod_{j=1}^{r_1} x_{1j}, \ \delta_2 = \prod_{j=1}^{r_2} x_{2j}.$

Integrating out $(\alpha, \theta_1, \theta_2)$ from (7.16) we have the joint posterior distribution of (p_1, p_2) is given by

$$\Pi(p_1, p_2 \mid \underline{x}) \propto p_1^{r_1} p_2^{r_2} \delta_1^{p_1} \delta_2^{p_2} \sum_{k=0}^{n-r} \binom{n-r}{k}$$

$$\left[\frac{B(n-k-r_2+1, r_2+k+1)}{\left\{ \sum_{j=1}^{r_2} x_{2j}^{p_2} + k \right\}^{r_2} \left\{ \sum_{j=1}^{r_1} x_{1j}^{p_1} + (n-r+k) \right\}^{r_1}} \right].$$

Integrating out p_2, the marginal posterior of p_1 is given by

$$\Pi(p_1 \mid \underline{x}) = C\, p_1^{r_1}\, \delta_1^{p_1} \sum_{k=0}^{n-r} \binom{n-r}{k} \cdot$$

$$\left[\frac{B(n-k-r_2+1,\, r_2+k+1)}{\left\{\sum_{j=1}^{r_1} x_{1j}^{p_1} + (n-r-k)\right\}^{r_1}} \int_0^\infty \frac{p_2^{r_2}\, \delta_2^{p_2}\, dp_2}{\left\{\sum_{j=1}^{r_2} x_{2j}^{p_2} + k\right\}^{r_2}}\right]. \tag{7.17}$$

Similarly,

$$\Pi(p_2 \mid \underline{x}) = C\, p_2^{r_2}\, \delta_2^{p_2} \sum_{k=0}^{n-r} \binom{n-r}{k} \cdot$$

$$\left[\frac{B(n-k-r_2+1,\, r_2+k+1)}{\left\{\sum_{j=1}^{r_2} x_{2j}^{p_2} + k\right\}^{r_2}} \int_0^\infty \frac{p_1^{r_1}\, \delta_1^{p_1}\, dp_1}{\left\{\sum_{j=1}^{r_1} x_{1j}^{p_1} + (n-r-k)\right\}^{r_1}}\right] \tag{7.18}$$

$$\Pi(\theta_1 \mid \underline{x}) = \frac{C}{\theta_1^{r_1+1}\,\Gamma(r_1)} \sum_{k=0}^{n-r} \binom{n-r}{k} \cdot$$

$$\left[B(n-k-r_2+1,\, r_2+k+1) \int_0^\infty \frac{p_2^{r_2}\, \delta_2^{p_2}\, dp_2}{\left\{\sum_{j=1}^{r_2} x_{2j}^{p_2} + k\right\}^{r_1}} \cdot\right.$$

$$\left.\int_0^\infty p_1^{r_1}\, \lambda_1^{p_1} \exp\left\{-\frac{T^{p_1}}{\theta_1}\left(\sum_{j=1}^{r_1} x_{1j}^{p_1} + n - r - k\right)\right\} dp_1\right]. \tag{7.19}$$

$$\Pi(\theta_2 \mid \underline{x}) = \frac{C}{\theta_2^{r_2+1}\,\Gamma(r_2)} \sum_{k=0}^{n-r} \binom{n-r}{k} \cdot$$

$$\left[B(n-k-r_2+1,\, r_2+k+1) \int_0^\infty \frac{p_1^{r_1}\, \delta_1^{p_1}\, dp_1}{\left\{\sum_{j=1}^{r_1} x_{1j}^{p_1} + (n-r-k)\right\}^{r_1}}\right.,$$

$$\int_0^\infty p_2^{r_2}\, \lambda_2^{p_2} \exp\left\{-\frac{T^{p_2}}{\theta_2}\left(\sum_{j=1}^{r_2} x_{2j}^{p_2}+k\right)\right\} dp_2\Bigg], \tag{7.20}$$

and,

$$\Pi(\alpha \mid \underline{x}) = C \sum_{k=0}^{n-r} \binom{n-r}{k} \cdot$$

$$\left[\alpha^{n-r_2-k}\,(1-\alpha)^{r_2+k} \int_0^\infty \frac{p_2^{r_2}\, \delta_2^{p_2}\, dp_2}{\left\{\sum_{j=1}^{r_1} x_{2j}^{p_2} + k\right\}^{r_2}},\right.$$

$$\left.\int_0^\infty \frac{p_1^{r_1}\, \delta_1^{p_1}\, dp_1}{\left\{\sum_{j=1}^{r_1} x_{1j}^{p_1} + (n-r-k)\right\}^{p_1}}\right] \tag{7.21}$$

where $\quad C^{-1} = \sum_{k=0}^{n-r} \binom{n-r}{k}\Big[B\,(n-k-r_2+1,\, r_2+k+1)\,.$

$$\left.\int_0^\infty \frac{p_2^{r_2}\, \delta_2^{p_2}\, dp_2}{\left\{\sum_{j=1}^{r_2} x_{2j}^{p_2} + k\right\}^{r_2}} \int_0^\infty \frac{p_1^{r_1}\, \lambda_1^{p_1}\, dp_1}{\left\{\sum_{j=1}^{r_1} x_{1j}^{p_1} + (n-r-k)\right\}^{r_1}}\right].$$

From the marginal posterior distributions (7.17) – (7.21) we obtain the corresponding Bayes estimators

$$p^t = C \sum_{k=0}^{n-r} \binom{n-r}{k}\Big[B\,(n-k-r_2+1,\, r_2+k+1)\,.$$

$$\left.\int_0^\infty \frac{p_1^{r_1+1}\, \delta_1^{p_1}\, dp_1}{\left\{\sum_{j=1}^{r_1} x_{1j}^{p_1}+(n-r-k)\right\}^{r_1}} \int_0^\infty \frac{p_2^{r_2}\, \delta_2^{p_2}\, dp_2}{\left\{\sum_{j=1}^{r_2} x_{2j}^{p_2}+k\right\}^{r_2}}\right]. \tag{7.22}$$

$$p_2^* = C \sum_{k=0}^{n-r} \binom{n-r}{k} \Bigg[B\,(n-k-r_2+1,\, r_2+k+1)\,.$$

$$\int_0^\infty \frac{p_1^{r_1}\, \delta_1^{p_1}\, dp_1}{\left\{ \sum\limits_{j=1}^{r_1} x_{1j}^{p_1} + (n-r-k) \right\}^{r_1}} \int_0^\infty \frac{p_2^{r_2+1}\, \delta_2^{p_2}\, dp_2}{\left\{ \sum\limits_{j=1}^{r_2} x_{2j}^{p_2} + k \right\}^{r_2}} \Bigg]. \tag{7.23}$$

$$\theta_1^* = \frac{C}{r_1-1} \sum_{k=0}^{n-r} \binom{n-r}{k} \Bigg[B\,(n-k-r_2,\, r_2+k+1)\,.$$

$$\int_0^\infty \frac{p_2^{r_2}\, \delta_2^{p_2}\, dp_2}{\left\{ \sum\limits_{j=1}^{r_2} x_{2j}^{p_2} + k \right\}^{r_2}} \int_0^\infty \frac{p_1^{r_1}\, (T\delta_1)^{p_1}\, dp_1}{\left\{ \sum\limits_{j=1}^{r_1} x_{1j}^{p_1} + (n-r-k) \right\}^{r_1-1}} \Bigg] \tag{7.24}$$

$$\theta_2^* = \frac{C}{r_2-1} \sum_{k=0}^{n-r} \binom{n-r}{k} \Bigg[B\,(n-k-r_2+1,\, r_2+k+1)\,.$$

$$\int_0^\infty \frac{p_2^{r_2}\, (T\delta_2)^{p_2}\, dp_2}{\left\{ \sum\limits_{j=1}^{r_2} x_{2j}^{p_2} + k \right\}^{r_2-1}} \int_0^\infty \frac{p_1^{r_1}\, (\delta_1)^{p_1}\, dp_1}{\left\{ \sum\limits_{j=1}^{r_1} x_{1j}^{p_1} + (n-r-k) \right\}^{r_1-1}} \Bigg] \tag{7.25}$$

$$\alpha^* = C \sum_{k=0}^{n-r} \binom{n-r}{k} \Bigg[B\,(n-r-k_2+2,\, r_2+k+1)\,.$$

$$\int_0^\infty \frac{p_2^{r_2}\, \delta_2^{p_2}\, dp_2}{\left\{ \sum\limits_{j=1}^{r_2} x_{2j}^{p_2} + k \right\}^{r_2}} \int_0^\infty \frac{p_1^{r_1}\, \delta_1^{p_1}\, dp_1}{\left\{ \sum\limits_{j=1}^{r_1} x_{1j}^{p_1} + (n-r-k) \right\}^{r_1}} \Bigg] \tag{7.26}$$

Unlike the mixture of exponentials, Bayes estimators (7.22) – (7.26) cannot be expressed in closed forms and one has to resort to some numerical integration routine. Given a data set one can compute the MLE (Sinha, 1982) and compare them with their Bayesian counterparts.

100 samples of different sizes and different sets of parameters were generated from the mixture distribution (7.9). It was observed that, (i) Bayes and MLE of (p_1, p_2, α) were almost identical; (ii) one of the two parameters θ_1 and θ_2 is over-estimated in both cases and the extent of over-estimation is greater in Bayesian than in the MLE; and (iii) Bayes estimators (θ^*_1, θ^*_2) are extremely sensitive to the estimators (p^*_1, p^*_2). (Sinha, 1987).

Sinha and Sloan (1989) have obtained Bayesian Predictive Intervals of a future observation from the mixture model (7.14).

Exercises

1. Show that for $r = n$, Bayes estimators of the parameters $(\theta^*_1, \theta^*_2, p^*)$ in the mixture model (7.1) and prior $g(\theta_1, \theta_2) \propto \dfrac{1}{\theta_1 \theta_2}$ are given by (7.9).

2. Consider the data in Example 7.2 truncated at $T = 100$.

   Obtain Bayes estimators of (θ_1, θ_2, p) under the prior $g(\theta_1, \theta_2, p) \propto \dfrac{1}{\theta_1 \theta_2}$.

3. Compute Bayes estimators (7.9) using the complete sample in Example 7.2.

4. Obtain Bayes estimators of the parameters $(\alpha, p_1, p_2, \theta_1, \theta_2)$ of the mixture of Weibull densities

$$f(t \mid \theta_1, \theta_2, p_1, p_2, \alpha) = \frac{\alpha}{\theta_1} t^{p_1 - 1} \exp\left(-\frac{t^{p_1}}{\theta_1}\right) + (1-\alpha)\, t^{p_2 - 1} \exp\left(-\frac{t^{p_2}}{\theta_2}\right)$$

$$p_1, p_2, \theta_1, \theta_2,\ t > 0, \quad 0 < \alpha < 1.$$

under the prior $g(\theta_1, \theta_2, p_1, p_2, \alpha) \propto \dfrac{1}{\theta_1 \theta_2}$ for the complete sample case $r = n$.

5. 100 units were exposed to a life testing experiment. Suppose the underlying failure-time distribution is a mixture of Weibulls as in Exercise 4 and the ordered failure times are recorded in units of 1000 hours. We observe $r_1 = 36$, $r_2 = 22$, $n = 100$ and the experiment is terminated at time $T = 2$. Compute Bayes estimates of $(\theta_1, \theta_2, p_1, p_2, \alpha)$ under the prior

$$g(\theta_1, \theta_2, p_1, p_2, \alpha) \propto \frac{1}{\theta_1 \theta_2}.$$

Sample from subpopulation 1 :

0.2027	0.5894	0.9511	1.2691	1.4022	1.6503
0.3784	0.6274	0.9554	1.2892	1.4213	1.7462
0.4556	0.6358	1.0471	1.3106	1.4444	1.7982
0.4896	0.7172	1.1007	1.3602	1.4684	1.8127
0.5565	0.8279	1.2212	1.3756	1.5682	1.8443
0.5719	0.9311	1.2434	1.3814	1.5951	1.8619

Sample from subpopulation 2 :

0.2994	0.9025	1.2593	1.3659	1.7251	1.9542
0.7655	1.0355	1.2622	1.4062	1.7491	1.9856
0.8518	1.0694	1.2839	1.5078	1.8200	
0.8842	1.1087	1.3535	1.5576	1.9359	

6. Use the data above and compute Bayes estimators of the parameters under the same prior for the complete sample case $n = 100$.

7. Let the following data represent a random sample of size $n = 25$ from the mixture distribution (7.1) truncated at $T = 100$.

Sample from the first subpopulation :

6.1886	8.7778	9.1949	9.7326	11.5059	13.0639	13.1145
26.3979	27.2176	28.7843	30.0532	31.0214	32.5078	36.1116
38.4326	46.7884	56.2416	91.1947	101.0766	110.0240	131.0561
208.2173						

Sample from the second subpopulation :

12.4904 49.0183 84.0717

Set up 95% prediction intervals similar to the Table in Example 7.2 for the same set of values of (a_i, b_i) and the same prior (7.11).

CHAPTER

8

Bayesian Break-even Analysis

8.1 Bayes' Rule

In Section 1 .3 we have stated and proved Bayes' Theorem

$$P(B_i \mid A) = \frac{P(B_i)\ P(A \mid B_i)}{\sum_{i=1}^{k} P(B_i)\ P(A \mid B_i)}. \tag{8.1}$$

We assume $P(B_i)$, the prior probability of the event B_i and the conditional probability $P(A \mid B_i)$ are known. Also known as Bayes' rule, (8,1) derives the probability of the event B_i, given the event A or assuming that the event A has already occurred.

Let $f(y \mid \theta)$ be the pdf of a random variable Y characterized by the parameter θ, and let $g(\theta)$ represent the prior distribution of θ. Then (8.1) may be written as

$$\Pi(\theta \mid y) = \frac{g(\theta)\ f(y \mid \theta)}{\int g(\theta)\ f(y \mid \theta)\ d\theta} \qquad -\infty < y < \infty, \quad -\infty < \theta < \infty.$$

$\Pi(\theta \mid y)$ is known as the posterior distribution of θ given the data y.

We will illustrate a simple application of the rule (8.1) with reference to the following business-related problem.

**Example 8.1:*

Women's clothing department of a large store has found, from past experience, that the probability that a customer will return a purchase is 0.10. It also has found that 70% of all purchases returned by customers are charged, and that 50% of all purchases not returned are charged.

(a) Use Bayes' Theorem to find the probability that if a purchase is charged it will be returned.

(b) Develop the prior probability distribution of a purchase return. Develop the posterior probability distribution of a purchase return, given that the purchase is charged.

(c) What is the probability that a cash purchase will not be returned?

(d) Suppose that 60% of the returned purchases are for \$15 or more and are charged, and that 45% of all purchases are for \$15 or more, and are charged. Find the probability that if a purchase is for \$15 or more and is charged, it will be returned.

(Neter, Wasserman and Whitmore, 1973).

Let A_1 be the event that a purchase is returned, B_1 be the event that a purchase is charged, and C_1 be the event that a purchase is for \$15 or more.

(a) From (8.1) we obtain

$$P(A_1/B_1) = \frac{P(A_1)\ P(B_1/A_1)}{P(A_1)\ P(B_1/A_1) + P(\bar{A}_1)\ P(B_1/\bar{A})}$$

$$= \frac{(0.10)\ (0.70)}{(0.10)\ (0.70) + (0.90)\ (0.50)}$$

$$= 0.135$$

(b) *Prior* *Posterior*

$P(A_1) = 0.10 \quad P(A_1/B_1) = 0.135$ ($\bar{A}$ represents the event

$P(\bar{A}_1) = 0.90 \quad P(\bar{A}_1/B_1) = 0.865$ not – A)

---

* With kind permission of publishers Allyn and Bacon.

(c) $$P(\bar{A}_1/\bar{B}_1) = \frac{P(\bar{A}_1)\ P(\bar{B}_1/\bar{A}_1)}{P(\bar{A}_1)\ P(\bar{B}_1/\bar{A}_1) + P(A_1)\ P(\bar{B}_1/A_1)}$$

$$= \frac{(0.90)\ (0.50)}{(0.90)\ (0.50) + (0.10)\ (0.30)}$$

$$= 0.9375$$

(d) $B_1 \cap C_1$ represents the event that the purchase is for \$15 or more and is charged.

$$P(A_1/B_1 \cap C_1) = \frac{P(A_1 \cap B_1 \cap C_1)}{P(B_1 \cap C_1)}$$

$$= \frac{P(A_1)\ P(B_1 \cap C_1/A_1)}{P(B_1 \cap C_1)}$$

$$= \frac{(0.10)\ (0.60)}{(0.45)}$$

$$= 0.1333.$$

8.2 Bayes' Act

A decision problem is usually characterized by the following parameters

(i) Acts available : $A = \{A_1, A_2, \ldots, A_m\}$

(ii) State of Nature: $\theta = \{\theta_1, \theta_2, \ldots, \theta_k\}$

(iii) Payoff: $X = \{X_1, X_2, \ldots, X_k\}$ corresponding to each act A_i, $i = 1, 2, \ldots m$.

The act that maximizes the expected payoff for each alternative decision is known as Bayes' Act.

(a) Decision based on prior probabilities alone.

Let $P(\theta)$ be the prior probability distribution of θ. Suppose the decision maker has no data ($n = 0$), and he has to make a decision about a proposed project. From past experience and his own expert opinion, he assigns *a-prior* probability distribution $P(\theta)$ for each payoff X as follows:

θ	$P(\theta)$	X
θ_1	P_1	X_1
θ_2	P_2	X_2
θ_3	P_3	X_3
.	.	.
.	.	.
θ_k	P_k	X_k.

$$E\text{ (payoff)} = \sum_{i=1}^{k} X_i P_i$$

Compute E (payoff) corresponding to each available act A_j. Bayes' Act is the act A_j such that E (payoff/A_j) is maximum.

Example 8.2:

A vendor is planning to buy a large lot of light bulbs from a wholesaler. The proportion θ of defective bulbs in the lot is unknown. From his past experience on the quality of similar products supplied by the wholesaler, he assigns prior probabilities $P(\theta)$ and figures out the corresponding profits (or payoff) X as follows:

θ	$P(\theta)$	X
0.02	0.1	\$300
0.04	0.2	100
0.06	0.3	0
0.08	0.2	- 50
0.10	0.2	- 100

Based on the available data, he has to make a decision if (A_1) he should accept the lot, or (A_2) reject it.

(a) Should the vendor buy the lot?

(b) Which act is Bayes' act?

Solution:

From the above table it follows that

$$E\text{ (payoff}/A_1) = \sum X P(\theta) = \$30 + 20 - 10 - 20) = \$20.00$$

$E\,(\text{payoff}/A_2) = \$0.00$

$E\,(\text{payoff}/A_1) > E\,(\text{payoff}/A_2)$.

The vendor should buy the lot.

Bayes' act is A_1.

(b) Revised degree of belief — decision based on prior probabilities and sample information.

Suppose we want to make an inference about a parameter θ. Any prior information we have about θ is reflected by the prior distribution $P(\theta)$. We now draw a sample from the underlying population. The sample gives us additional information about θ. In the Bayesian framework, we combine the prior information and the sample information (B) and obtain a revised degree of belief about θ or the posterior distribution of θ given by $P(\theta/B)$ using Bayes' rule (8.1).

Example 8.3:

Refer to Example 8.2. Suppose a random sample of 20 bulbs was drawn from the lot and 2 bulbs were found defective. In the light of this new information.

(a) Which act is Bayes' Act?

(b) Should the lot be accepted?

Solution:

Let r be the number of defectives in a sample of n items. r is binomially distributed with parameters (n, θ).

$$P(r) = \binom{n}{r} \theta^r (1-\theta)^{n-r}, \quad r = 0, 1, 2, \ldots, n.$$

Let B_1 represent the situation $n = 20$, $r = 2$. We write

$B_1 : n = 20,\ r = 2.$

$$P(r = 2 \mid \theta = 0.02) = \binom{20}{2} (0.02)^2 (0.98)^{18} = 0.05283, \text{ and similarly}$$

for other θ.

θ	$P(\theta)$	$P(B_1/\theta)$	$P(\theta)P(B_1/\theta)$	$P(\theta/B_1)$
0.02	0.1	0.0528	0.00528	0.0248
0.04	0.2	0.1458	0.02916	0.1368
0.06	0.3	0.2246	0.06738	0.3162
0.08	0.2	0.2711	0.05422	0.2545
0.10	0.2	0.2852	0.05704	0.2677
	1.0		$0.21308 = P(B_1)$	1.000

$$E(\text{payoff}/A_1) = \sum X_i P(\theta/B_1)$$

$$= -\$18.375$$

$$E(\text{payoff}/A_2) = \$0.00 > E(\text{payoff}/A_1)$$

Bayes Act is A_2.

Do not accept the lot.

Example 8.4:

Suppose a second sample of 15 bulbs were drawn from the lot and there were no defectives in the sample.

On the basis of the posterior distribution, $P(\theta \mid B_1)$ in Example 8.3 and the new information

$$B_2 : n = 15, \ r = 0.$$

which one is Bayes' act?

Solution:

We will now use $P(\theta \mid B_1)$ as our new prior distribution $P(\theta)$.

θ	$P(\theta)$	$P(B_2 \mid \theta)$	$P(\theta)P(B_2 \mid \theta)$	$P(\theta \mid B_2)$
0.02	0.0248	0.7386	0.01832	0.0530
0.04	0.1368	0.5421	0.07416	0.2147
0.06	0.3162	0.3953	0.12499	0.3618
0.08	0.2545	0.2863	0.07286	0.2109
0.10	0.2677	0.2059	0.05512	0.1596
	1.0000		$P(B_2) = 0.34545$	1.0000

$$E\,(\text{payoff}/A_1) = \$10.865$$

$$E\,(\text{payoff}/A_2) = \$0.00 < E\,(\text{payoff}/A_1)$$

Bayes Act is A_1

Accept the lot.

Example 8.5:

Suppose we pool the two samples and the number of defectives in the pooled sample. Based on the prior distribution $P(\theta)$ and the sample information,

$$B : n = 35,\ \ r = 2,$$

which act is Bayes Act?

Solution :

$$P(r) = \binom{35}{r}\ \theta^r\ (1-\theta)^{35-r}.$$

$$P(r = 2 \mid \theta = 0.02) = \binom{35}{2}\ (0.02)^2\ (0.98)^{33}$$

$$= 0.1222$$

θ	$P(\theta)$	$P(B \mid \theta)$	$P(\theta)\,P(B \mid \theta)$	$P(\theta \mid B)$
0.02	0.1	0.1222	0.01222	0.0530
0.04	0.2	0.2475	0.04950	0.2148
0.06	0.3	0.2780	0.08340	0.3618
0.08	0.2	0.2430	0.04860	0.2108
0.10	0.2	0.1839	0.03678	0.1596
	1.0		$P(B) = 0.2305$	1.0000

Same posterior is obtained as in Example 8.4 and, hence, the same Bayes' Act: A_1.

Accept the lot.

We will obtain the same posterior whether the priors are successively revised or whether the several small samples are pooled into a large sample and the initial prior is used to derive the posterior.

*Example 8.6 :

A chemical plant has two major fractionating columns which require occasional internal cleaning and maintenance. During the mid-year maintenance period, the plant manager must decide to either strip both columns for cleaning and maintenance (C_1) or leave the columns untouched until the end of the year (C_2) when cleaning and maintenance of both columns will be undertaken in either case. The condition of the two columns — that is, whether or not they require cleaning and maintenance — is the key determinant of the payoffs in this decision problem. The payoffs (in thousands of dollars) for the two acts and the different possible outcome states are shown below. It is impossible to accurately assess the condition of the columns by external inspection alone. However, on the basis of personal judgement and experience, the plant manager is able to assign the prior probabilities shown below to the different outcome states:

Number requiring cleaning and maintenance	Prior Probability	Payoff of Act	
		C_1	C_2
0	0.6	30	40
1	0.1	20	20
2	0.3	10	0

(a) Suppose the plant manager must choose between C_1 and C_2 without obtaining further information on the condition of the two columns. Which act is the Bayes' Act?

(b) Suppose, on the other hand, that the plant manager has the option of having one or both columns opened and inspected before he makes his decision. He arbitrarily decides to have one of the two columns inspected and picks one column at random. The inspection shows that this column requires cleaning and maintenance. Revise the plant manager's prior probabilities on the basis of this sample information by means of Bayes' Theorem.

(c) Identify the Baye's Act for the sample result in part (*b*).

(d) Determine the Bayes' decision rule when one column is inspected ($n = 1$).

---

* With kind permission of publishers Allyn and Bacon.

(e) Suppose the cost of having a column opened and inspected is \$1000. What is the expected payoff net of sampling costs (the cost of opening and inspecting the column) for the Bayes' decision rule when $n = 1$?

(f) Determine the expected payoff net of sampling costs for the Bayes' decision rule based on $n = 2$. Compare this expected net payoff with those of the Bayes' decision rules for $n = 0$ in part (a) and for $n = 1$ in part (e). What is the optimal sample size here?

(Neter, Wasserman and Whitmore, 1973).

Solution:

Let θ be the number of columns requiring cleaning and maintenance, $P(\theta)$ be the prior probability distribution of θ and X_1, X_2 be the payoffs corresponding to the acts C_1 and C_2, respectively.

θ	$P(\theta)$	X_1	X_2
0	0.6	30	40
1	0.1	20	20
2	0.3	10	0

(a) $E(\text{payoff}/C_1) = 30 \times .6 + 20 \times .1 + 10 \times .3 = \$23$

$E(\text{payoff}/C_2) = 40 \times .6 + 20 \times .1 = \$26 > E(\text{payoff}/C_1)$

C_2 is Bayes' Act.

(b) Let $B_1 : n = 1,\ \theta = 1$

represent the event that one of the two columns is chosen at random and the inspection shows that it needs cleaning and maintenance.

θ	$P(\theta)$	$P(B_1/\theta)$	$P(\theta)\,P(B_1/\theta)$	$P(\theta/B_1)$
0	0.6	0	0	0
1	0.1	0.5	0.05	$\frac{5}{35} = 0.143$
2	$\underline{0.3}$	1	$\underline{0.30}$	$\frac{30}{35} = 0.857$
	1.0		$P(B_1) = 0.35$	1.0

On the basis of the sample information B_1, the revised probability distribution of θ is given by $P(\theta / B_1)$ tabulated above.

(c) $E(\text{payoff}/C_1) = 0.143 \times 20 + 0.857 \times 10 = \$11.43$

$E(\text{payoff}/C_2) = 0.143 \times 20 = \$2.86 < E(\text{payoff}/C_1)$.

Hence, Bayes' Act is C_1.

(d) Given $n = 1$, we may have $\theta = 1$ or $\theta = 0$. We have discussed the event $B_1 : n = 1, \theta = 1$ in (b). Consider the event $B_2 : n = 1, \theta = 0$.

θ	$P(\theta)$	$P(B_2/\theta)$	$P(\theta)\,P(B_2/\theta)$	$P(\theta/B_2)$
0	0.6	1	0.60	$\frac{60}{65} = 0.923$
1	0.1	0.5	0.05	$\frac{5}{65} = 0.077$
2	0.3	0	0.00	0.000
	1.0		$P(B_2) = 0.65$	1.0

$E(\text{payoff}/C_1) = 0.923 \times 30 + 0.077 \times 20 = \$29.23$

$E(\text{payoff}/C_2) = 0.923 \times 40 + 0.077 \times 20 = \$38.46$

Bayes' Act is C_2.

(e) We observe that

if $\theta = 0$, Bayes' Act is C_2, payoff $= \$38.46$;

if $\theta = 1$, Bayes' Act is C_1, payoff $= \$11.43$.

Hence, Bayes' decision rule corresponding to $n = 1$ is:

Choose C_2 if $\theta = 0$.

Choose C_1 if $\theta = 1$.

(f) $B_i \quad P(B_i) \quad E(\text{payoff}/B_i)$

$$n = 1 \begin{cases} B_1 & 0.35 & 11.43 \\ B_2 & 0.65 & 38.46 \end{cases}$$

Sampling Cost $= \$1000$.

$E(\text{net payoff}) = (11.43 \times 0.35 + 38.46 \times 0.65 - 1) \times 1000$

$$= \$\ 27.9995 \times 1000$$

$$= \$\ 28{,}000.$$

(g) When both the columns are opened and inspected, the plant manager has the perfect information — that is, the exact information about the condition of the columns. Once he has the perfect information, his optimal acts will be to choose

C_1	if $\theta = 0$
C_1 or C_2	if $\theta = 1$ and
C_1	if $\theta = 2$.

With the perfect information we now have the following distribution of θ, (θ) and payoff X_i.

θ	$P(\theta)$	X
0	0.6	40
1	0.1	20
2	0.3	10

$$E\,(\text{payoff / perfect information}) = 40 \times 0.6 + 20 \times 0.1 + 10 \times 0.3 = \$\ 29$$

$$E\,(\text{net payoff / perfect information}) = \$\ 1{,}000\,(29 - 2) = \$\ 27{,}000$$

n	E (net payoff)
0	\$26,000
1	28,000
2	27,000

In the subsequent sections we will be using the following terms:

(i) Opportunity Loss (OL)

For a given state of nature let A_i be the act that maximizes the profit $P(\theta, A_i)$ and let

$$P_i = \max\ P(\theta,\ A_i).$$

Let A_j be another act that does not maximize $P(\theta,\ A_j)$.

The opportunity loss (OL) is defined as the difference between the profit associated with the optimal act A_i and any

other act A_j. Thus, $OL(\theta, A_j) = P_i - P(\theta, A_j)$ opportunity loss is also known as Regret.

Let $l(\theta, A)$ represent the loss associated with the state θ and Act A.

If loss is interpreted as being negative profit, we define

$$R(\theta, A_j) = l(\theta, A_j) - \min l(\theta, A_j). \tag{8.2}$$

For example, suppose we have the following six available acts for three states of nature and the associated 'losses' $l(\theta, A)$ are:

Act / θ	A_1	A_2	A_3	A_4	A_5	A_6
θ_1	2	6	3	5	3	3
θ_2	3	7	4	6	4	2
θ_3	1	4	2	3	3	2

Min $l(\theta_1, A_i) = 2$, achieved for A_1

Min $l(\theta_2, A_i) = 2$, achieved for A_6

Min $l(\theta_3, A_i) = 1$, achieved for A_1

Using (8.2) we set up the Regret table

$R(\theta_i, A_i)$

Act / θ	A_1	A_2	A_3	A_4	A_5	A_6
θ_1	0	4	1	3	1	1
θ_2	1	5	2	4	2	0
θ_3	0	3	1	2	2	1
Max $R(\theta_i, A_i)$	1	5	2	4	2	1

If *min* (*max*. regret) is used as a criterion for optimal act, A_1 and A_6 are both optimal.

(ii) Expected opportunity loss (*EOL*).

Let $P(\theta)$ be the prior probability distribution of θ and $L(\theta)$ be the corresponding opportunity loss. The expected opportunity loss is defined by

$$EOL = \Sigma_\theta \; P(\theta) \; L(\theta).$$

If no data is available, the act that minimizes the EOL is the optimal act.

Suppose the probability distribution of θ in the preceding example is given by

$$P(\theta_1) = \frac{1}{2}, \; P(\theta_2) = \frac{1}{4}, \; P(\theta_3) = \frac{1}{4}.$$

$$[EOL/A_i] = \frac{1}{4}, \frac{16}{4}, \frac{5}{4}, \frac{12}{4}, \frac{6}{4}, \frac{4}{3}, \text{ respectively}$$

$$Min \; [EOL/A_i] = \frac{1}{4} \text{ for } A_1 \text{ which is optimal.}$$

Given the data D

$$EOL = \Sigma_\theta \; L(\theta) \; P(\theta/D).$$

(iii) Expected value of the perfect information ($EVPI$).

If we had the perfect information about the state θ, the EOL should be zero. Thus, EOL may be regarded as a measure of the cost of uncertainty (Jedamus, Frame and Taylor, 1976), and as such, EOL is also called $EVPI$.

*Example 8.7 :

A company is trying to decide what size plant to build in a certain area. Three alternatives are being considered: plants with capacities of 10,000, 15,000 and 20,000 units respectively. Demand for the product is uncertain, but management has assigned the probabilities listed below to five levels of demand. The payoff table below also shows the profit (in millions of dollars) for each alternative and each possible level of demand (output may exceed rated capacity).

* With kind permission of publishers Richard Irwin.

Demand in units Z	Probability $P(Z)$	Actions: build plant with capacity of		
		10, 000 units	15, 000 units	20, 000 units
5, 000	0.2	−4.0	−6.0	−8.0
10, 000	0.3	1.0	0.0	−2.0
15, 000	0.2	1.5	6.0	5.0
20, 000	0.2	2.0	7.5	11.0
25, 000	0.1	2.0	8.0	12.0

(a) What size plant should be built?

(b) Prepare an opportunity loss table.

(c) What is *EVPI*? (Spurr and Bonini, 1973).

(a) Expected payoff under 10,000 units

$$= -4 \times .2 + .3 \times 1 + .2 \times 1.5 + 2 \times .2 + 1. \times 2 = .4 .$$

Expected payoff under 15,000 units

$$= -6 \times .2 + 6 \times .2 + 7.5 \times .2 + 8 \times .1 = 2.3 .$$

Expected payoff under 20,000 units

$$= -8 \times .2 - 2 \times .3 + 5 \times .2 + 11 \times .2 + 12 \times .1 = 2.2 .$$

Expected payoff is maximum under 15,000 units. Hence, the plant with 15,000 units should be built.

(b)

Demand Z	Max. payoff
5, 000	−4.0
10, 000	1.0
15, 000	6.0
20, 000	11.0
25, 000	12.0

Opportunity Loss/Regret Table:

Z	P (Z)	10, 000	15, 000	20, 000
5, 000	0.2	0	2.0	4.0
10, 000	0.3	0	1.0	3.0
15, 000	0.2	4.5	0	1.0
20, 000	0.2	9.0	3.5	0
25, 000	0.1	10.0	4.0	0
EOL		3.7	1.8	1.9

(c) $EOL/10{,}000$ units

$$= .2 \times 4.5 + .2 \times 9.0 + .1 \times 10$$

$$= 3.7, \text{ and}$$

similarly, for expected payoff under 15,000 and 20,000 units.

(d) From (*b*) it follows that *EOL* is minimum under 15,000 units, i.e., if we build the plant with capacity 15,000 units we will have less regret for the opportunity lost than if we build the plant with any other capacity.

Thus, $EVPI = 1.8$

8.3 Break-even Analysis

As the term suggests, the break-even point is a state where the company or a business does not incur any loss nor does it make any profit. In a business decision problem it is important that the firm has some knowledge about the break-even point. This information may be obtained using the prior knowledge alone ($n = 0$) or prior knowledge as well as the sample information ($n = n_0$).

(a) Without sampling ($n = 0$).

Consider the following problem:

Example 8.8:

Suppose a firm selling to 20,000 industrial accounts is considering adding a new product to its line. Management is somewhat reluctant to make this proposed addition because its development costs are high. It is estimated that $500,000 would be required for full-scale development. Further, the management believes the proposed product to be a one-shot items, having virtually no possibility of repeat sales to the same customer. It is unique, so its addition can be expected to have no effect, either positive or negative, on sales of the existing line.

The product would add nothing to the existing overhead expenses (above the $500,000 development costs) and could be produced at a

* With kind permission of publishers McGraw-Hill.

variable cost of $1200 per unit. Variable selling expenses would be approximately $300 per unit. Because of the nature of the item, it can be produced to order, so no inventory need be carried.

As a result of previous experience with similar products and interviews with company salesmen, the market research department believes the product could be most profitably priced at $2000. They are able to assign the probability distribution in the following table to the proportion of customers who will actually purchase the new product.

Probability distribution for new product

Proportion purchasing: p	Probability: $P(p)$
0.02	0.1
0.04	0.3
0.06	0.3
0.08	0.2
0.10	0.1
	1.0

(8.3)

(i) Compute the economic break-even value of p.

(ii) Compute the COL (conditional opportunity loss) for the decision 'develop' and 'do not develop'.

(iii) Compute the EOL of an immediate decision without sampling.

(iv) What is the correct decision? (Jedamus, Frame and Taylor, 1976).

Solution:

(i) We have net profit/unit

= selling price – (production cost + marketing expenses)

= $2000 – $(1200 + 300)

= $500.

Let N be the number of customers that must buy the product for the firm to break even.

Fixed cost = $500,000

Total profit generated by sales to N customers = $500 N

For break-even state,

$$500\,N = 500{,}000$$

$$N = 1000.$$

$$\text{Break-even proportion } (BEP) = \frac{1000}{20{,}000} = 0.05.$$

If the break-even proportion $p < 0.05$ and the firm goes ahead with the project, they will loss money; if $p > 0.05$ and the firm does not develop the product, an opportunity loss will occur and the company will lose money. If $p \geq 0.05$, and the company develops the product or if $p < 0.05$ and the company shelves the project, there will not be any opportunity.

Since the opportunity loss is conditional on the level of p, we will call it conditional opportunity loss (COL). If $p < 0.05$ and the act is "DEVELOP", the company has to pay \$500,000 towards the first developing cost. A part of this cost will be reimbursed by the profit \$500 per item sold to an unknown proportion p of 20,000 customers.

(ii) Hence, the COL = $[500{,}000 - (500)\,(20{,}000)\,p]$

$$= \$\,(500.000 - 10{,}000{,}000)\,p]$$

$$= \$\,10{,}000{,}000\,(0.05 - p).$$

If $p < 0.05$ and the product is not developed, the $COL = 0$. Similarly, if $p > 0.05$ and the act is "DEVELOP", the $COL = 0$; on the other hand, if $p > 0.05$ and the act is "DO NOT DEVELOP",

$$\text{the COL} = \$(10{,}000{,}000\,p - 500{,}000)$$

$$= \$10{,}000{,}000\,(p - 0.05).$$

We construct the COL – functions in the following

COL – functions

	Develop	Do not develop
$p < 0.05$	$10, 000, 000 (0.05 – p)	0
$p > 0.05$	0	$10, 000, 000 ($p$ – 0.05)

(8.4)

(iii) Using the probability distribution of p in (8.3) and the COL – functions above, we construct the $EOL = \sum_p P(p)\ COL$.

Event		Act			
		Develop		Do not develop	
p	$P(p)$	COL	$P(p)\ COL$	COL	$P(p)\ COL$
0.02	0.1	\$300, 000	\$30, 000	0	0
0.04	0.3	100, 000	30, 000	0	0
0.06	0.3	0	0	\$100,000	\$30, 000
0.08	0.2	0	0	300,000	60, 000
0.10	0.1	0	0	500,000	50, 000
EOL			\$60, 000		\$140, 000

(8.5)

(iv) The (EOL/develop) < (EOL/do not develop).

The optional act would be "DEVELOP" the product.

(b) With sampling ($n = n_0$).

Given $n = 0$, from (8.4) it follows that the decisions (Develop/$p <$ 0.05) and (Do not develop/$p >$ 0.05) for the specified decision rule D: ($n = n_0$, $c = c_0$) are wrong decisions. Now suppose the company has decided that they will select a random sample of $n = n_0$ customers and if at least c number of customers agree to buy the product, they will develop it otherwise the project will be cancelled. One may now compute the P (wrong decision $\cap\ p$) and the corresponding EOL.

$$= \sum_p (COL)\ (\text{Probability of wrong decision } \cap\ p)$$

for the specified decision rule,

$$D : (n = n_0,\ c = c_0)$$

For a given n, the EOL may be computed (as above) for different sets of c and c^* will be the optimal criteria number which minimizes the $[EOL]_{c = c^*}$.

*Example 8.9 :

Consider the Example 8.8 and suppose the company decides on selecting a random sample of 20 customers, and if at least one customer agrees to buy the product, they will develop, it.

(Jedamus, Frame and Taylor, 1976).

We have Y = no. of customers who agree to buy the product

$$\sim B(n, p).$$

Note that ("Develop / $p < 0.05$") implies a wrong decision.

$$P(\text{wrong devision} / p = 0.02,\ n = 20)$$
$$= P(r \geq 1 \mid p = 0.02,\ n = 20)$$
$$= 1 - P(r = 0 \mid p = 0.02,\ n = 20)$$
$$= 1 - (0.98)^{20} = 0.3324$$

Similarly, $P(\text{wrong decision} \mid p = 0.04,\ n = 20) = 0.5580.$

"Develop / $p > 0.05$" implies the correct decision.

Hence, $P(\text{wrong decision} / p = 0.06,\ n = 20)$

$$= 1 - P(r \geq 1 \mid p = 0.06,\ n = 20)$$
$$= P(r = 0 \mid p = 0.06,\ n = 20)$$
$$= (0.94)^{20} = 0.2901$$

and similarly for $p = 0.08, 0.10$.

$$P(\text{wrong decision} \cap p) = P(p)\, P(\text{wrong decision} / p).$$

Thus, $P(\text{wrong decision} \cap p = 0.02) = (0.1)(0.3324) = 0.03324.$

Let D_1 be the event: $n = 20$, $c = 1$.

$$\therefore\ P(\text{wrong decision}) = \Sigma_p\, P(\text{wrong decision} \cap p)$$
$$= \Sigma_p\, P(p)\, P(\text{wrong decision} / p).$$

Probability of wrong decision / D_1 (Table 20.6, Jedamus, Frame and Taylor, 1976).

P	$P(p)$	P (wrong decision / p)	P (wrong decision $\cap p$)	
0.02	0.1	0.3324	0.03324	
0.04	0.3	0.5580	0.16740	
0.06	0.3	0.2901	0.08703	
0.08	0.2	0.1887	0.03774	
0.10	0.1	0.1216	0.01216	
			0.33757	$= P$ (wrong decision)

Marginal probability of making a wrong decision = 0.33757, i.e. the probability that the decision rule $D_1 : n = 20, c = 1$ will lead to a wrong decision is approximately 1 in 3.

$$(EOL / D_1) = \sum_p \{Prob \text{ (wrong decision } \cap p)\} \bullet \{COL / \text{wrong decision}\}$$

The (COL / wrong decision) are easily obtained from (8.5).

(EOL / D_1) : (Jedamus, Frame and Taylor, 1973, Table 20.7)

(1) P	(2) COL / Wrong decision	(3) P (wrong decision $\cap p$)	4 = (2) × (3)
0.02	\$300, 000	0.03324	\$9, 972
0.04	100, 000	0.16740	16, 740
0.06	100, 000	0.08703	8, 703
0.08	300, 000	0.03774	11, 322
0.10	500, 000	0.01216	6, 080
EOL			\$52, 817

(c) Optimal criterion number c^*.

Consider the Example 8.8 and the decision to develop the product if $D_2 : n = 20, c = 2$ is attained

P (wrong decision $/ p = 0.02, n = 20$)

$$= p\,(r \geq 2 \mid p = 0.02,\ n = 20)$$

$$= 1 - \sum_{0}^{1} \binom{20}{r} (0.02)^r \ (0.98)^{20-r}$$

$$= 1 - 0.9401$$

$$= 0.0599$$

Similarly, $P(\text{wrong decision} \mid p = 0.04,\ n = 20) = 0.1897$

P (wrong decision $/ p = 0.06,\ n = 20)$

$$= P(r \le 1 \mid p = 0.06,\ n = 20)$$

$$= \sum_{r=0}^{1} \binom{20}{r} (0.06)^r\ (0.94)^{20-r}$$

$$= 0.6605.$$

Similarly, for $p = 0.08,\ 0.10$.

Probability (wrong decision/D_2) (Jedamus, Frame and Taylor, 1973, Table 20.8).

p	$P(p)$	P (wrong decision / p)	P (wrong decision $\cap\ p$)
0.02	0.1	0.0599	0.00599
0.04	0.3	0.1897	0.05691
0.06	0.3	0.6605	0.19815
0.08	0.2	0.5169	0.10338
0.10	0.1	0.3917	0.03917
P (wrong decision)			0.40360

Proceeding as in EOL/D_1, it is easy to show that EOL/D_2 = \$77,902. Thus, EOL for the decision rule D_2: $n = 20,\ c = 2 >$ the EOL for the decision rule D_1 : $n = 20,\ c = 1$, and the corresponding EOL for other criterion number $c = 3, 4, \ldots$ (given the same $n = 20$) will lead to progressively worse results. Hence, given $n = 20$, $c^* = 1$ is the optimal criterion number.

(Jedamus, Frame and Taylor, 1976).

Exercises

1. A florist buys 4 dozen long stem roses per day from a vendor at the cost of \$10/dozen and sells them for \$30/dozen. No order for single or partial dozen is entertained. Suppose the

probability distribution of X = dozen of roses sold/day is given by

X	0	1	2	3	4
$P(X)$	0.2	0.3	0.3	0.1	0.1

At the end of the day, unsold roses (if any) are sent out to hospitals for free.

(a) What is the expected profit if 4 dozen roses are purchased?

(b) What is the expected profit if the florist purchases,

(i) 3 dozen

(ii) 2 dozen?

(c) How many dozen roses should the florist keep in stock to make the most profit?

*2. A company is considering the introduction of a new product. The planned selling price is \$410 per unit. There is a fixed cost of \$200,000 for developing and manufacturing the product, while the variable cost would be \$210 per unit. Demand for the product is described by the following distribution.

Demand	Probability
1000	0.05
1500	0.10
2000	0.15
2500	0.50
3000	0.20

(a) Based on the current available data, should the company introduce the product?

(b) What is the probability of a loss?

(c) What is the *EVPI*? (Forgionne, 1986)

*3. JM Enterprises, Inc. is going to buy a new machine that could generate significant labor cost savings on a certain operation. The company is considering two brands, X and Y. Brand X involves a fixed charge of \$500,000 and a variable operating

cost of $250 per hour of operation. Brand Y has a fixed expense of $600,000 and a variable operating cost of $200 per hour. Annual operating hours are distributed as follows:

Operating Hours	Probability
100	0.1
200	0.2
300	0.3
400	0.2
500	0.2

(a) Based on a current available data, which brand should JM Enterprises buy?

(b) Develop the operating loss function.

(c) What is the *EVPI*? (Forgionne, 1986).

*4. The personnel director of a steel company has learned from past experience that the probability is 0.80 that a management trainee successfully completes the two-year program. The proportion of trainees with previous business experience at another company is 0.10 among the trainees successfully completing the program, and 0.20 among those who do not successfully complete the program.

(a) Use Bayes' Theorem to find the probability that if a trainee has previous business experience he will successfully complete the training program.

(b) Develop both the prior probability distribution of training success and the posterior probability distribution given that the trainee has previous business experience.

(c) Do either the prior or posterior probabilities take account of the trainee's academic record? Discuss.

(d) Are the differences between the prior and posterior probability distributions necessarily due to the effect of previous business experience? Would information about persons not hired for the management training

program be relevant for answering this question? Discuss.

(Neter, Wasserman, Whitmore, 1973).

*5. Refer to Example 8.6

(a) Prior to opening and inspecting a column, what is the plant manager's expected payoff with perfect information?

(b) Compare the expected payoff in part (*a*) with the expected payoff net of sampling costs for the Bayes' decision on $n = 2$ in part (*f*) of Example 8.6. What accounts for the difference in these two values?

(c) What is the expected value of perfect information to the plant manager? How can perfect information be obtained in this problem?

(Neter, Wasserman, Whitmore, 1973).

*6. Refer to the product-development problem in Example 8.9.

Sow that the *EOL* for the decision rule $D_2 : n = 20,\ c = 2$ is \$77,902.

(Jedamus, Taylor and Frame, 1976).

**7. Consider the following payoff table.

		Actions				
Event	Probability	*A*	*B*	*C*	*D*	*E*
I	0.05	100	120	210	140	180
II	0.05	110	160	190	140	180
III	0.10	130	200	170	140	100
IV	0.30	150	180	120	140	180
V	0.40	180	150	100	140	120
VI	0.10	250	100	100	140	120

(a) Prepare an opportunity loss table for this decision situation.

(b) Calculate the expected opportunity loss for each action.

(c) What is *EVPI*?

(Spurr and Bonini, 1973).

8. Refer to the product development problem in Example 8.8.

   Suppose the fixed cost is \$600,000, production cost is \$1300/unit, variable selling expenses are \$400/unit, the selling price is \$2500/unit and the company has 15000 customers. Suppose the probability distribution of customers who will buy the product is given by:

Proportion Purchasing	Probability
0.02	0.2
0.04	0.2
0.06	0.3
0.08	0.2
0.10	0.1

   (a) Compute the break-even proportion p.

   (b) Compute the *EOL* for the decision—develop/do not develop.

   (c) What is the correct decision?

9. Refer to the Exercise #8.

   Suppose the management decided to draw a random sample of 15 customers and if at least one customer wants to buy the product, they would develop it.

   What is the probability that the decision rule $D : (n = 15, c = 1)$ will lead to a wrong decision?

*10. The Theta Company manufactures its requirements for part No. 805 in lots of 1000 units. It has been difficult to control the quality of this product without a complicated readjustment of the manufacturing equipment. The cost of such a redjustment is \$300. When such a readjustment has been made, only 2 percent defectives are produced. Without the adjustment, the quality has been quite variable, as shown by the history of the last 20 lots:

---

* With kind permission of publisher Richard Irwin.

Percent defectives	No. of lots
0.02	5
0.05	8
0.10	4
0.15	2
0.20	1
	20

A lot of part No. 805 is about to be manufactured, and management is undecided about whether it should pay for the cost by adjustment or take the chance of a large percentage of defectives. Replacement cost of a defective item is $5.

(a) Draw up a payoff table and calculate the expected cost of each action using the past frequency data as prior probabilities. Which action is preferable?

(b) What is the *EVPI*?

(c) Suppose the manufacturing process was set up and the first 20 items were examined and 2 defectives were found. Should the machine be shut down and an adjustment made at this time, or should the should the manufacturing process be allowed to continue?

(Spurr and Bonini, 1973).

11. (Continuation of Problem 10).

Suppose the sample result had been no defectives out of 20 items sampled. What is the expected posterior loss of each action Which action is preferable? What is the posterior *EVPI*?

(Spurr and Bonini, 1973).

*12. (Continuation of Problems 10 and 11).

(a) Find the expected posterior costs for other relevant sample results.

(b) Suppose it costs $20 plus $2 per items sampled. Should a sample of 20 items be taken?

(Spurr and Bonini, 1973).

* With kind permission of publisher Richard Irwin.

APPENDIX

A.1

$$\int_0^\infty \frac{\exp\left(-\frac{a}{\theta}\right)}{\theta_{b+1}}\, d\theta = \frac{\Gamma(b)}{a^b}, \quad a, b > 0.$$

Put $\quad \frac{a}{\theta} = u, \; \theta = \frac{a}{u}, \; d\theta = -\frac{adu}{u^2}$

$$\int_0^\infty \frac{\exp\left(-\frac{a}{\theta}\right)}{\theta^{b+1}}\, d\theta = \int_0^\infty \left(\frac{u}{a}\right)^{b+1} \frac{a}{u^2} \exp(-u)\; du$$

$$= \frac{1}{a^b} \int_0^\infty u^{b-1} \exp(-u)\; du$$

$$= \frac{\Gamma(b)}{a^b}.$$

A. 2

$$\int_{-\infty}^\infty \exp(-ax^2)\; dx = \frac{\sqrt{\pi}}{\sqrt{a}}$$

Put $\quad x = \sqrt{\frac{u}{a}}, \; dx = \frac{1}{2\sqrt{a}}\, u^{(1/2)-1}\; du$

$$\int_{-\infty}^\infty \exp(-ax^2)\; dx = 2\int_0^\infty \exp(-ax^2)\; dx$$

$$= 2\int_0^\infty \frac{1}{2\sqrt{a}}\, u^{(1/2)-1} \exp(-u)\; du$$

$$= \frac{1}{\sqrt{a}} \int_0^{\infty} u^{\frac{1}{2}-1} \exp(-u)\, du = \frac{\Gamma\left(\frac{1}{2}\right)}{\sqrt{a}} = \sqrt{\frac{\pi}{a}}$$

A. 3

$$\frac{S}{\theta} \sim \gamma(n) \text{ where}$$

$$f(x \mid \theta) = \frac{1}{\theta} \exp\left(-\frac{x}{\theta}\right), \quad x, \theta > 0 \quad (1)$$

$$S = \sum_{i=1}^{n} x_i .$$

Put $\frac{x}{\theta} = u, \; dx = \theta\, du .$

From (1) we have

$$g(u) = \exp(-u), \quad 0 < u < \infty .$$

Thus, $$\frac{X}{\theta} = U \sim \gamma(1) \ldots \quad (2)$$

Let X and Y be independent gamma variates with parameters l and m. The joint distribution of (X, Y) is given by

$$f(x,y) = \frac{1}{\Gamma(l)\,\Gamma(m)} x^{l-1} y^{m-1} \exp\{-(x+y)\} .$$

Put $$\left.\begin{aligned} x + y &= u \\ y &= v \end{aligned}\right\} \begin{aligned} x &= x - v \\ y &= v . \end{aligned}$$

The Jacobian $|J| = \left|\frac{\partial(x, y)}{\partial(u, v)}\right| = \begin{vmatrix} 1 & -1 \\ 0 & 1 \end{vmatrix} = 1.$

The joint distribution of U and v is given by

$$g(u, v) = \frac{1}{\Gamma(l)\,\Gamma(m)} \exp(-u)\,(u-v)^{l-1}\, v^{m-1}, \quad 0 < v < u\, \infty$$

$g_1(u)$ = marginal distribution of u

$$= \int_0^{u} g(u, v)\, dv$$

$$= \frac{1}{\Gamma(l)\,\Gamma(m)} \exp(-u) \int_0^u (u-v)^{l-1}\, v^{m-1}\, dv$$

But $\quad v = u\,\theta$. For fixed u,

$$dv = u\, d\theta.$$

$$g_1(u) = \frac{1}{\Gamma(l)\,\Gamma(m)} \exp(-u)$$

$$\int_0^1 (u-u\theta)^{l-1}\,(u-\theta)^{m-1}\, u\, d\theta$$

$$= \frac{u^{l+m-1}}{\Gamma(l)\,\Gamma(m)} \exp(-u) \int_0^1 \theta^{m-1}\,(1-\theta)^{l-1}\, d\theta$$

$$= \frac{u^{l+m-1}\exp(-u)\, B(m,l)}{\Gamma(l)\,\Gamma(m)}$$

$$= \frac{u^{l+m-1}\exp(-u)}{\Gamma(l+m)},\quad u>0$$

which implies that $X + Y = U \sim$ as a gamma-variate with parameter $(l+m)$. (3)

Going back to (2).

$$\frac{X_i}{\theta} \sim \gamma(1).$$

Since $\left(\frac{X_i}{\theta}, \frac{X_2}{\theta}, \ldots, \frac{X_n}{\theta}\right)$ are independent gamma-variates each with parameter 1, from (3) it follows that

$$\frac{\Sigma X_i}{\theta} \sim \gamma(n),\quad i = 1, 2, \ldots, n.$$

or $$\frac{S}{\theta} \sim \gamma(n).$$

A.4

$$\sum_{i=0}^{k-1} \frac{(-1)^i \binom{k-1}{i}}{m-k+1+i} = B(k, m-k+1).$$

By definition,

$$B(a, b) = \int_0^1 x^{a-1} (1-x)^{b-1} \, dx$$

$$= \sum_{i=0}^{b-1} (-1)^i \binom{b-1}{i} \int_0^1 x^{a+i-1} \, dx$$

$$= \sum_{i=0}^{b-1} \frac{(-1)^i \binom{b-1}{i}}{a+i}.$$

Put $\qquad a = m - k + 1, b = k$

$$\sum_{i=0}^{b-1} \frac{(-1)^i \binom{-k-1}{i}}{m-k+1+i} = B(m-k+1, k) = B(k, m-k+1).$$

A.5

$$k \binom{m}{k} \sum_{i=0}^{k-1} \frac{(-1)^i \binom{k-1}{i}}{m-k+1+n+i} = \frac{m^{(k)}}{(m+n)^{(k)}}$$

where $m^{(k)} = m(m-1)(m-2) \ldots (m-k+1)$.

Replacing m by $(m+n)$ in A.4, we have

$$\sum_{i=0}^{k-1} \frac{(-1)^i \binom{k-1}{i}}{m+n-k+1+i} = B(k, m+n-k+1).$$

Multiplying both sides by $k \binom{m}{k}$,

$$k \binom{m}{k} \sum_{i=0}^{k-1} \frac{(-1)^i \binom{k-1}{i}}{m+n-k+1+i} = k \binom{m}{k} B(k, m+n-k+1)$$

$$= \frac{k\, m!}{k!\,(m-k)!} \frac{\Gamma(k)\, \Gamma(m+n-k+1)}{\Gamma(m+n+1)}$$

$$= \frac{k\, m!\, (k-1)!\, (m+n-k)!}{k!\, (m-k)\,!\, (m+n)!}$$

$$= \frac{m\,(m-1)\,(m-2)\,\ldots\,(m-k+1)}{(m+n)\,(m+n-1)\,(m+n-2)\,\ldots\,(m+n-k+1)}$$

$$= \frac{m^{(k)}}{(m+n)^{(k)}}.$$

A.6

$h(\theta) =$ Hartigan-prior for θ

$\propto \frac{1}{\theta^3},$ where

$$f(x \mid \theta) = \frac{x}{\theta^2} \exp\left(-\frac{x^2}{2\theta^2}\right), \quad x, \theta > 0. \tag{4}$$

We have

$$\log f = \log x - 2 \log \theta - \frac{x^2}{2\theta^2}$$

$$l_1 = \frac{\partial}{\partial\theta} \log f = \frac{-2}{\theta} + \frac{x^2}{\theta^3}$$

$$l_2 = \frac{\partial^2}{\partial\theta^2} \log f = \frac{2}{\theta^2} - \frac{3x^2}{\theta^4}$$

$$l_1^2 + l_2 = \frac{6}{\theta_2} - \frac{7x^2}{\theta^4} + \frac{x^4}{\theta^6}$$

$$l_1\, l_2 = \frac{-4}{\theta^3} + \frac{8x^2}{\theta^5} - \frac{3x^4}{\theta^7}.$$

Note that if $X \sim \gamma(l)$

$$E(X) = \text{Var}(X) = l.$$

From (4) it follows that

$$\frac{X^2}{2\theta^2} \sim \gamma(1)$$

$$E\left(\frac{X^2}{2\theta^2}\right) = 1, \quad E(X^2) = 2\theta^2$$

$$\text{Var}\left(\frac{X^2}{2\theta^2}\right) = 1, \quad \text{Var}(X^2) = 4\theta^2$$

$$E(X^4) = \text{Var}(X^2) + \{E(X^2)\}^2$$

$$= 8\,\theta^4.$$

$$E(l_1) = -\frac{2}{\theta} + \frac{2}{\theta} = 0$$

$$E(l_1^2 + l_2) = \frac{6}{\theta^2} - \frac{14}{\theta^2} + \frac{8}{\theta^2} = 0,$$

Thus, $\quad E(l_1) = E(l_1^2 + l_2) = 0.$

$$E(l_1\, l_2) = -\frac{4}{\theta^3} + \frac{16}{\theta^3} - \frac{24}{\theta^3} = -\frac{12}{\theta^3}$$

$$E(l_2) = \frac{2}{\theta^2} - \frac{6}{\theta^2} = -\frac{4}{\theta^2}$$

$$\frac{\partial}{\partial\theta}\log\{h(\theta)\} = -\frac{E(l_1\, l_2)}{E(l_2)} = -\frac{3}{\theta}$$

$$\log\{h(\theta)\} = -3\log\theta = \log\left(\frac{1}{\theta^3}\right)$$

or $\quad h(\theta) \propto \dfrac{1}{\theta^3}.$

A.7

$$B(a, b) = \int_0^\infty \frac{x^{a-1}}{(1+x)^{a+b}}\, dx.$$

By definition

$$B(a, b) = \int_0^\infty u^{a-1}\,(1-u)^{b-1}\, du.$$

Put $\quad u = \dfrac{1}{1+x}, \quad du = -\dfrac{1}{(1+x)^2}$

$$B(a, b) = \int_0^1 u^{a-1}\,(1-u)^{b-1}\, du$$

$$= \int_0^\infty \left(\frac{1}{1+x}\right)^{a-1} \left(\frac{x}{1+x}\right)^{b-1} \frac{1}{(1+x)^2}\, dx$$

$$= \int_0^\infty \frac{x^{b-1}}{(1+x)^{a+b}}\, dx .$$

Interchanging a and b

$$B(b, a) = \int_0^\infty \frac{x^{a-1}}{(1+x)^{a+b}}\, dx .$$

Since $\quad B(a, b) = B(b, a)$

$$B(a, b) = \int_0^\infty \frac{x^{a-1}}{(1+x)^{a+b}}\, dx .$$

A.8

Elements of the $[-L_{ij}]$ matrix.

From (6.23)

$$L = \text{Constant} - n\left[\log \sigma_1 + \log \sigma_2 + \frac{1}{2}\log(1-\rho^2)\right] \cdot$$

$$-\frac{1}{2(1-\rho^2)} \sum_{j=1}^{n} \left[\frac{(x_{1j}-\mu_1)^2}{\sigma_1^2} + \frac{(x_{2j}-\mu_2)^2}{\sigma_2^2} - \frac{2\rho}{\sigma_1 \sigma_2}(x_{1j}-\mu_1)(x_{2j}-\mu_2)\right]$$

$$L_1 = \frac{\partial L}{\partial \mu_1} = \frac{1}{1-\rho^2} \sum_{j=1}^{n} \left[\frac{x_{1j}-\mu_1}{\sigma_1^2} - \frac{\rho}{\sigma_1 \sigma_2}(x_{2j}-\mu_2)\right]$$

$$L_{11} = \frac{\partial 2\, L}{\partial \mu_1^2} = \frac{1}{1-\rho^2} \sum_{j=1}^{n} \left(-\frac{1}{\sigma_1^2}\right) = \frac{-n}{\sigma_1^2 (1-\rho^2)}$$

$$L_{12} = \frac{\partial^2 L}{\partial \mu_1\, \partial \mu_2} = \frac{\rho}{1-\rho^2} \sum_{j=1}^{n} \frac{1}{\sigma_1 \sigma_2} = \frac{n\rho}{(1-\rho^2)\,\sigma_1 \sigma_2}$$

$$L_{13} = \frac{\partial^2 L}{\partial \mu_1\, \partial \sigma_1} = \frac{1}{1-\rho^2} \sum_{j=1}^{n} \left[-\frac{2(x_{1j}-\mu_1)}{\sigma_1^3} + \frac{\rho}{\sigma_1^2 \sigma_2}(x_{2j}-\mu_2)\right]$$

$$= 0 \text{ at } \hat{\mu}_1 = \bar{x}_1,\ \hat{\mu}_2 = \bar{x}_2 .$$

Similarly,

$$L_{14} = \frac{\partial^2 L}{\partial \mu_1\, \partial \sigma_2} = 0 \text{ and}$$

$$L_{15} = \frac{\partial^2 L}{\partial \mu_1\, \partial \rho} = 0 \text{ at } \hat{\theta} = (\hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}_1, \hat{\sigma}_2, \hat{\rho})\,.$$

$$L_2 = \frac{\partial L}{\partial \mu_2} = \frac{1}{1-\rho^2} \sum_{j=1}^{n} \left[\frac{x_{2j} - \mu_2}{\sigma_2^2} - \frac{\rho\,(x_{1j} - \mu_1)}{\sigma_1\, \sigma_2} \right]$$

$$L_{21} = \frac{\partial^2 L}{\partial \mu_2\, \partial \mu_1} = \frac{n\rho}{\sigma_1\, \sigma_2\, (1-\rho^2)}$$

$$L_{22} = \frac{\partial^2 L}{\partial \mu_2^2} = \frac{1}{1-\rho^2} \sum_{j=1}^{n} \left(-\frac{1}{\sigma_2^2} \right)$$

$$= -\frac{n}{\sigma_2^2\,(1-\rho^2)}$$

$$L_{23} = \frac{\partial^2 L}{\partial \mu_2\, \partial \sigma_1}\,, \quad L_{24} = \frac{\partial^2 L}{\partial \mu_2\, \partial \sigma_2}$$

and $\quad L_{25} = \dfrac{\partial^2 L}{\partial \mu_2\, \partial \rho}$ all vanish at $\hat{\theta}$.

$$L_3 = \frac{\partial L}{\partial \sigma_1} = -\frac{n}{\sigma_1} + \frac{1}{1-\rho^2}\,.$$

$$\sum_{j=1}^{n} \left[\frac{(x_{1j} - \mu_1)^2}{\sigma_1^3} - \frac{\rho}{\sigma_1^2\, \sigma_2} (x_{1j} - \mu_1)\,(x_{2j} - \mu_2) \right]$$

$$L_{31} = \frac{\partial^2 L}{\partial \sigma_1\, \partial \mu_1} = \frac{1}{1-\rho^2}\,.$$

$$\sum_{j=1}^{n} \left[-\frac{2\,(x_{1j} - \mu_1)}{\sigma_1^3} + \frac{\rho}{\sigma_1^2\, \sigma_2} (x_{2j} - \mu_2) \right]$$

$$= 0 \text{ at } \hat{\mu}_1 = \bar{x}_1,\ \hat{\mu}_2 = \bar{x}_2\,.$$

$$L_{32} = \frac{\partial^2 L}{\partial \sigma_1 \, \partial \mu_2} = \frac{1}{1-\rho^2} \sum_{j=1}^{n} \frac{\rho}{\sigma_1^2 \sigma_2} (x_{1j} - \mu_1)$$

$$= 0 \text{ at } \hat{\mu}_1 = \bar{x}_1 .$$

$$L_{33} = \frac{\partial^2 L}{\partial \sigma_1^2} = \frac{n}{\sigma_1^2} + \frac{1}{1-\rho^2} \cdot$$

$$\sum_{j=1}^{n} \left[-\frac{3 (x_{1j} - \mu_1)^2}{\sigma_1^4} + \frac{2\rho}{\sigma_1^3 \sigma_2} (x_{1j} - \mu_1)(x_{2j} - \mu_2) \right]$$

$$\left[L_{33}\right]_\theta = \frac{n}{\sigma_1^2} + \frac{1}{1-\rho^2} \left[-\frac{3n}{\sigma_1^2} + \frac{2n\rho^2}{\sigma_1^2} \right]$$

$$= \frac{n}{\sigma_1^2 (1-\rho^2)} [1 - \rho^2 - 3 + 2\rho^2]$$

$$= -\frac{n (2-\rho^2)}{\sigma_1^2 (1-\rho^2)} .$$

Similarly,

$$\left[L_{34}\right]_\theta = \left[\frac{\partial^2 L}{\partial \sigma_1 \, \partial \sigma_2}\right]_\theta = -\frac{n (2-\rho^2)}{\sigma_2^2 (1-\rho^2)}$$

$$L_{35} = \frac{\partial^2 L}{\partial \sigma_1 \, \partial \rho} = \frac{2\rho}{(1-\rho^2)^2} \cdot$$

$$\sum_{j=1}^{n} \left[\frac{(x_{1j} - \mu_1)^2}{\sigma_1^3} - \frac{\rho}{\sigma_1^2 \sigma_2} (x_{1j} - \mu_1)(x_{2j} - \mu_2) \right] -$$

$$\frac{1}{1-\rho^2} \sum_{j=1}^{n} \left[\frac{(x_{1j} - \mu_1)(x_{2j} - \mu_2)}{\sigma_1^2 \sigma_2} \right]$$

$$\left[L_{35}\right]_\theta = \frac{2\rho}{(1-\rho^2)^2} \left[\frac{n}{\sigma_1} - \frac{n\rho^2}{\sigma_1} \right] - \frac{np}{\sigma_1 (1-\rho^2)}$$

$$= \frac{2\rho n - n\rho}{\sigma_1 (1-\rho^2)} = \frac{n\rho}{\sigma_1 (1-\rho^2)}$$

$$L_4 = \frac{\partial L}{\partial \sigma_2} = -\frac{n}{\sigma_2} + \frac{1}{1-\rho^2}.$$

$$\sum_{j=1}^{n} \left[\frac{(x_{2j}-\mu_2)^2}{\sigma_2^3} - \frac{\rho}{\sigma_1 \sigma_2^2}(x_{1j}-\mu_1)(x_{2j}-\mu_2)\right]$$

$$L_{41} = \frac{\partial^2 L}{\partial \sigma_2 \partial \mu_1} = \frac{1}{1-\rho^2} \sum_{j=1}^{n} \frac{\rho}{\sigma_1 \sigma_2^2}(x_{2j}-\mu_2)$$

$$= \frac{\rho}{(1-\rho^2)\sigma_1 \sigma_2^2} \sum_{j=1}^{n} (x_{2j}-\mu_2) = 0 \text{ at } \hat{\mu}_2 = \bar{x}_2.$$

$$L_{42} = \frac{\partial^2 L}{\partial \sigma_2 \partial \mu_2} = \frac{1}{1-\rho^2} \sum_{j=1}^{n} \left[-\frac{2(x_{2j}-\mu_2)}{\sigma_2^3} + \frac{\rho(x_{1j}-\mu_1)}{\sigma_1 \sigma_2^2}\right]$$

$$= 0 \text{ at } \hat{\mu}_1 = \bar{x}_1,\ \hat{\mu}_2 = \bar{x}_2.$$

$$L_{43} = \frac{\partial^2 L}{\partial \sigma_2 \partial \sigma_1} = \frac{1}{1-\rho^2} \sum_{j=1}^{n} \frac{\rho(x_{1j}-\mu_1)(x_{2j}-\mu_2)}{\sigma_1^2 \sigma_2^2}$$

$$= \frac{n\rho^2}{\sigma_1 \sigma_2 (1-\rho^2)} \text{ at } \hat{\mu}_1 = \bar{x}_1,\ \hat{\mu}_2 = \bar{x}_2.$$

$$L_{44} = \frac{\partial^2 L}{\partial \sigma_2^2} = \frac{n}{\sigma_2^2} + \frac{1}{1-\rho^2}.$$

$$\sum_{j=1}^{n} \left[-\frac{3(x_{2j}-\mu_2)^2}{\sigma_2^4} + \frac{2\rho}{\sigma_1 \sigma_2^3}(x_{1j}-\mu_1)(x_{2j}-\mu_2)\right]$$

$$\left[L_{44}\right]_\theta = \frac{n}{\sigma_2^2} + \frac{1}{1-\rho^2}\left[\frac{-3n}{\sigma_2^2} + \frac{2n\rho^2}{\sigma_2^2}\right] = \frac{-n(2-\rho^2)}{\sigma_2^2(1-\rho^2)}$$

$$L_{45} = \frac{\partial^2 L}{\partial \sigma_2 \partial \rho} = \frac{2\rho}{(1-\rho^2)^2}.$$

$$\sum_{j=1}^{n} \left[\frac{(x_{2j}-\mu_2)^2}{\sigma_2^3} - \frac{\rho}{\sigma_1 \sigma_2^2}(x_{1j}-\mu_1)(x_{2j}-\mu_2)\right]$$

$$- \frac{1}{(1-\rho^2)\,\sigma_1\,\sigma_2^2} \sum_{j=1}^{n} (x_{1j}-\mu_1)(x_{2j}-\mu_2).$$

$$\left[L_{45}\right]_\theta = \frac{2\rho}{(1-\rho^2)^2}\left(\frac{n}{\sigma_2}-\frac{n\rho^2}{\sigma_2}\right) - \frac{n\rho}{(1-\rho^2)\,\sigma_2} = \frac{n\rho}{\sigma_2\,(1-\rho^2)}$$

$$L_5 = \frac{\partial L}{\partial \rho} = \frac{n\rho}{(1-\rho^2)} - \frac{\rho}{(1-\rho^2)^2}.$$

$$\sum_{j=1}^{n}\left[\frac{(x_{1j}-\mu_1)^2}{\sigma_1^2} + \frac{(x_{2j}-\mu_2)^2}{\sigma_2^2} - \frac{2\rho}{\sigma_1\,\sigma_2}(x_{1j}-\mu_1)(x_{2j}-\mu_2)\right]$$

$$+ \frac{1}{(1-\rho^2)\,\sigma_1\,\sigma_2} \sum_{j=1}^{n} (x_{1j}-\mu_1)(x_{2j}-\mu_2)$$

$$L_{51} = \frac{\partial^2 L}{\partial \rho\, \partial \mu_1} = -\frac{\rho}{(1-\rho^2)^2}.$$

$$\sum_{j=1}^{n}\left[-\frac{2\,(x_{1j}-\mu_1)}{\sigma_1^2} + \frac{2\rho}{\sigma_1\,\sigma_2}(x_{2j}-\mu_2)\right]$$

$$- \frac{1}{(1-\rho^2)\,\sigma_1\,\sigma_2} \sum_{j=1}^{n} (x_{2j} - \mu_2)$$

$$= 0 \text{ at } \hat{\mu}_1 = \bar{x}_1,\ \hat{\mu}_2 = \bar{x}_2.$$

$$L_{52} = \frac{\partial^2 L}{\partial \rho\, \partial \mu_2} = -\frac{\rho}{(1-\rho^2)^2}.$$

$$\sum_{j=1}^{n}\left[-\frac{2\,(x_{2j}-\mu_2)}{\sigma_2^2} + \frac{2\rho}{\sigma_1\,\sigma_2}(x_{1j}-\mu_1)\right]$$

$$- \frac{1}{(1-\rho^2)\,\sigma_1\,\sigma_2} \sum_{j=1}^{n} (x_{1j}-\mu_1)$$

$$= 0 \text{ at } \hat{\mu}_1 = \bar{x}_1,\ \hat{\mu}_2 = \bar{x}_2.$$

$$L_{53} = \frac{\partial^2 L}{\partial \rho \, \partial \sigma_1} = -\frac{\rho}{(1-\rho^2)^2} \sum_{j=1}^{n} \Bigg[$$

$$\left[-\frac{2\,(x_{1j}-\mu_1)^2}{\sigma_1^3} + \frac{2\rho\,(x_{1j}-\mu_1)\,(x_{2j}-\mu_2)}{\sigma_1^2\,\sigma_2}\right]\Bigg]$$

$$-\frac{1}{(1-\rho^2)\,\sigma_1^2\,\sigma_2} \sum_{j=1}^{n} (x_{1j}-\mu_1)\,(x_{2j}-\mu_2)$$

$$\left[L_{53}\right]_\theta = -\frac{\rho}{(1-\rho^2)^2}\left(-\frac{2n}{\sigma_1} + \frac{2n\rho^2}{\sigma_1}\right) - \frac{n\rho}{(1-\rho^2)\,\sigma_1} = \frac{n\rho}{(1-\rho^2)\,\sigma_1}.$$

Similarly,

$$\left[L_{54}\right]_\theta = \left[\frac{\partial^2 L}{\partial \rho \, \partial \sigma_2}\right]_\theta = \frac{n\rho}{(1-\rho^2)\,\sigma_2}$$

$$L_{55} = \frac{\partial^2 L}{\partial \rho^2} = \frac{n\,(1+\rho^2)}{(1-\rho^2)^2} - \frac{1+3\rho^2}{(1-\rho^2)^3}.$$

$$\sum_{j=1}^{n}\left[\frac{(x_{1j}-\mu_1)^2}{\sigma_1^2} + \frac{(x_{2j}-\mu_2)^2}{\sigma_2^2} - \frac{2\rho\,(x_{1j}-\mu_1)\,(x_2 j-\mu_2)}{\sigma_1\,\sigma_2}\right]$$

$$+\frac{4\rho}{\sigma_1\,\sigma_2\,(1-\rho^2)^2} \sum_{j=1}^{n} \left(x_{1j}-\mu_1)\,(x_{2j}-\mu_2\right)$$

$$\left[L_{55}\right]_\theta = \frac{n\,(1+\rho^2)}{(1-\rho^2)^2} - \frac{1+3\rho^2}{(1-\rho^2)^3}\,(n+n-2n\rho^2) + \frac{4n\rho^2}{(1-\rho^2)^2}$$

$$= \frac{n\,(1+\rho^2)}{(1-\rho^2)^2} - \frac{2n\,(1+3\rho^2)}{(1-\rho^2)^2}$$

$$+ \frac{4n\rho^2}{(1-\rho^2)^2} = -\frac{n\,(1+\rho^2)}{(1-\rho^2)^2}.$$

A.9

L_{ijk} Terms

The bivariate normal log-likelihood function is

$$L(\underline{x} \mid \underline{\theta}) = -n \log(2\pi) - \frac{n}{2}\left\{\log \sigma_1^2 + \log \sigma_2^2 + \log(1-\rho^2)\right\}$$

$$- \frac{1}{2(1-\rho^2)} \sum_{j=1}^{n} \left\{ \left(\frac{x_{1j}-\mu_1}{\sigma_1} \right)^2 + \left(\frac{x_{2j}-\mu_2}{\sigma_2} \right)^2 \right.$$

$$\left. - 2\rho \left(\frac{x_{1j}-\mu_1}{\sigma_1} \right) \left(\frac{x_{2j}-\mu_2}{\sigma_2} \right) \right\}.$$

The third partial derivative terms (L_{ijk}) are:

$$L_{111} = \frac{\partial^3 L}{\partial \mu_1^3} = 0$$

$$L_{112} = \frac{\partial^3 L}{\partial \mu_1^2 \partial \mu_2} = 0$$

$$L_{113} = \frac{\partial^3 L}{\partial \mu_1^2 \partial \sigma_1} = \frac{2n}{\sigma_1^3 (1-\rho^2)}$$

$$L_{114} = \frac{\partial^3 L}{\partial \mu_1^2 \partial \sigma_2} = 0$$

$$L_{115} = \frac{\partial^3 L}{\partial \mu_1^2 \partial \rho} = \frac{-2n\rho}{\sigma_1^2 (1-\rho^2)^2}$$

$$L_{122} = \frac{\partial^3 L}{\partial \mu_1 \partial \mu_2^2} = 0$$

$$L_{123} = \frac{\partial^3 L}{\partial \mu_1 \partial \mu_{22} \partial \sigma_1} = \frac{-n\rho}{\sigma_1^2 \sigma_2 (1-\rho^2)}$$

$$L_{124} = \frac{\partial^3 L}{\partial \mu_1 \partial \mu_2 \partial \sigma_2} = \frac{-n\rho}{\sigma_1 \sigma_2^2 (1-\rho^2)}$$

$$L_{125} = \frac{\partial^3 L}{\partial \mu_1 \partial \mu_2 \partial \rho} = \frac{n(1+\rho^2)}{\sigma_1 \sigma_2 (1-\rho^2)^2}$$

$$L_{133} = \frac{\partial^3 L}{\partial \mu_1 \partial \sigma_1^2}$$

$$= \frac{1}{\sigma_1^2 (1-\rho^2)} \left[\frac{6\Sigma (x_{1j}-\mu_1)}{\sigma_1^2} - \frac{2\rho \, \Sigma (x_{2j}-\mu_2)}{\sigma_1 \sigma_2} \right] = 0 \text{ at } \hat{\theta}.$$

$$L_{134} = \frac{\partial^3 L}{\partial \mu_1\, \partial \sigma_1\, \partial \sigma_2} = \frac{-\rho\, \Sigma\,(x_{2j} - \mu_2)}{\sigma_1^2\, \sigma_2^2\, (1 - \rho^2)} = 0 \text{ at } \hat{\theta}\,.$$

$$L_{135} = \frac{\partial^3 L}{\partial \mu_1\, \partial \sigma_1\, \partial \rho} = \frac{-4\rho\, \Sigma\,(x_{1j} - \mu_1)}{\sigma_1^3\, (1 - \rho^2)^2} + \frac{(1 + \rho^2)\, \Sigma\,(x_{2j} - \mu_2)}{(1 - \rho^2)^2\, \sigma_1^2\, \sigma_2}$$

$$= 0 \text{ at } \hat{\theta}\,.$$

$$L_{144} = \frac{\partial^3 L}{\partial \mu_1\, \partial \sigma_2^2} = \frac{-2n\rho\, (\bar{x}_2 - \mu_2)}{\sigma_1\, \sigma_2^3\, (1 - \rho^2)} = 0 \text{ at } \hat{\theta}\,.$$

$$L_{145} = \frac{\partial^3 L}{\partial \mu_1\, \partial \sigma_2\, \partial \rho} = \frac{n\,(1 + \rho^2)\,(\bar{x}_2 - \mu_2)}{\sigma_1\, \sigma_2^2\, (1 - \rho^2)^2} = 0 \text{ at } \hat{\theta}$$

$$L_{155} = \frac{2n}{(1 - \rho^2)^3\, \sigma_1} \left[\frac{(\bar{x}_1 - \mu_1)(1 + 3\rho^2)}{\sigma_1} - \frac{\rho\,(3 + \rho^2)\,(\bar{x}_2 - \mu_2)}{\sigma_2} \right] = 0 \text{ at } \hat{\theta}\,.$$

$$L_{222} = \frac{\partial^3 L}{\partial \mu_2^3} = 0$$

$$L_{223} = \frac{\partial^3 L}{\partial \mu_2^2\, \partial \sigma_1} = 0$$

$$L_{224} = \frac{\partial^3 L}{\partial \mu_2^2\, \partial \sigma_2} = \frac{2n}{\sigma_2^3\, (1 - \rho^2)}$$

$$L_{225} = \frac{\partial^3 L}{\partial \mu_2^2\, \partial \rho} = \frac{-2n\rho}{\sigma_2^2\, (1 - \rho^2)^2}$$

$$L_{233} = \frac{\partial^3 L}{\partial \mu_2\, \partial \sigma_1^2} = \frac{-2n\rho\, (\bar{x}_1 - \mu_1)}{\sigma_1^3\, \sigma_2\, (1 - \rho^2)^2} = 0 \text{ at } \hat{\theta}\,.$$

$$L_{234} = \frac{\partial^3 L}{\partial \mu_2\, \partial \sigma_1\, \partial \sigma_2} = \frac{-n\rho\, (\bar{x}_1 - \mu_1)}{\sigma_1^2\, \sigma_2^2\, (1 - \rho^2)} = 0 \text{ at } \hat{\theta}$$

$$L_{235} = \frac{\partial^3 L}{\partial \mu_2\, \partial \sigma_1\, \partial \rho} = \frac{n\,(1 + \rho^2)\,(\bar{x}_1 - \mu_1)}{\sigma_1^2\, \sigma_2\, (1 - \rho^2)^2} = 0 \text{ at } \hat{\theta}$$

$$L_{244} = \frac{\partial^3 L}{\partial \mu_2\, \partial \sigma_2^2} = \frac{1}{\sigma_2^2 (1-\rho^2)} \left[\frac{6n(\bar{x}_2 - \mu_2)}{\sigma_2^2} - \frac{2n\rho(\bar{x}_1 - \mu_1)}{\sigma_1 \sigma_2} \right]$$

$$= 0 \text{ at } \hat{\theta}.$$

$$L_{245} = \frac{\partial^3 L}{\partial \mu_2\, \partial \sigma_2\, \partial \rho}$$

$$= \frac{1}{\sigma_2^2 (1-\rho^2)^2} \left[\frac{n(1+\rho^2)(\bar{x}_1 - \mu_1)}{\sigma_1} - \frac{4n\rho(\bar{x}_2 - \mu_2)}{\sigma_2} \right] = 0 \text{ at } \hat{\theta}.$$

$$L_{255} = \frac{\partial^3 L}{\partial \mu_2\, \partial \rho^2}$$

$$= \frac{2n}{\sigma_2 (1-\rho^2)^3} \left[\frac{(3\rho^2+1)(\bar{x}_2 - \mu_2)}{\sigma_2} - \frac{\rho(3+\rho^2)(\bar{x}_1 - \mu_1)}{\sigma_1} \right]$$

$$= 0 \text{ at } \hat{\theta}.$$

$$L_{333} = \frac{\partial^3 L}{\partial \sigma_1^3} = \frac{-2}{\sigma_1^3 (1-\rho^2)}.$$

$$\left[n(1-\rho^2) - \frac{6 \sum_{j=1}^{n} (x_{1j} - \mu_1)^2}{\sigma_1^2} + \frac{3\rho\, \Sigma (x_{1j} - \mu_1)(x_{2j} - \mu_2)}{\sigma_1 \sigma_2} \right]$$

$$L_{334} = \frac{\partial^3 L}{\partial \sigma_1^2\, \partial \sigma_2} = \frac{-2\rho \Sigma (x_{1j} - \mu_1)(x_{2j} - \mu_2)}{\sigma_1^3 \sigma_2^2 (1-\rho^2)}$$

$$L_{335} = \frac{\partial^3 L}{\partial \sigma_1^2\, \partial \rho}$$

$$= \frac{2(1+\rho^2)\, \Sigma (x_{11j} - \mu_1)(x_{2j} - \mu_2)}{\sigma_1^3 \sigma_2 (1-\rho^2)} - \frac{6\rho\, \Sigma (x_{1j} - \mu_1)^2}{\sigma_1^4 (1-\rho^2)^2}$$

$$L_{344} = \frac{\partial^3 L}{\partial \sigma_1\, \partial \sigma_2^2} = \frac{-2\rho \sum_{j=1}^{n} (x_{1j} - \mu_1)(x_{2j} - \mu_2)}{\sigma_1^2 \sigma_2^3 (1-\rho^2)}$$

$$L_{345} = \frac{\partial^3 L}{\partial \sigma_1\, \partial \sigma_2\, \partial \rho} = \frac{(1+\rho^2)\, \Sigma (x_{1j} - \mu_1)(x_{2j} - \mu_2)}{\sigma_1^2 \sigma_2^2 (1-\rho^2)^2}$$

$$L_{355} = \frac{\partial^3 L}{\partial \sigma_1 \, \partial \rho^2} = \frac{2}{\sigma_1^2 (1-\rho^2)^3} \cdot$$

$$\left[\frac{(1+3\rho^2)\,\Sigma\,(x_{1j}-\mu_1)^2}{\sigma_1} - \frac{(3\rho+\rho^3)\,\Sigma\,(x_{1j}-\mu_1)(x_{2j}-\mu_2)}{\sigma_2}\right]$$

$$L_{444} = \frac{\partial^3 L}{\partial \sigma_2^3} = \frac{-2}{\sigma_2^3 (1-\rho^2)} \cdot$$

$$\left[n(1-\rho^2) - \frac{6 \sum_{j=1}^{n} (x_{2j}-\mu_2)^2}{\sigma_2^2} + \frac{3\rho\,\Sigma\,(x_{1j}-\mu_1)(x_{2j}-\mu_2)}{\sigma_1 \sigma_2}\right]$$

$$L_{445} = \frac{\partial^3 L}{\partial \sigma_2^2 \, \partial \rho} = \frac{2(1+\rho^2)\,\Sigma\,(x_{1j}-\mu_1)(x_{2j}-\mu_2)}{\sigma_1 \sigma_2^3 (1-\rho^2)^2} - \frac{6\rho\,\Sigma\,(x_{2j}-\mu_2)}{\sigma_2^4\,(1-\rho^2)^2}$$

$$L_{455} = \frac{\partial^3 L}{\partial \sigma_2 \, \partial \rho^2} = \frac{2}{\sigma_2^2 (1-\rho^2)^3} \cdot$$

$$\left[\frac{(1+3\rho^2)\,\Sigma\,(x_{2j}-\mu_2)^2}{\sigma_2} - \frac{(3\rho+\rho^3)\,\Sigma\,(x_{1j}-\mu_1)(x_{2j}-\mu_2)}{\sigma_1}\right]$$

$$L_{555} = \frac{\partial^3 L}{\partial \rho^3} = \frac{2}{(1-\rho^2)^3} \cdot$$

$$\left[n\rho(\rho^2+3) - \frac{6\rho(\rho^2+1)}{(1-\rho^2)} \left(\frac{\Sigma\,(x_{1j}-\mu_1)^2}{\sigma_1^2} + \frac{\Sigma\,(x_{2j}-\mu_2)^2}{\sigma_2^2}\right)\right.$$

$$\left. + \frac{3(1+6\rho^2+\rho^4)}{(1-\rho^2)} \left(\frac{\Sigma\,(x_{1j}-\mu_1)(x_{2j}-\mu_2)}{\sigma_1 \sigma_2}\right)\right]$$

Note: $L_{ijk} = L_{ikj} = L_{jik} = L_{jki} = L_{kij} = L_{kji} \; \forall_{i,j,k}$ for all (i, j, k)

A.10

BVN $(\mu_1, \mu_2\ \sigma^2, \sigma^2, \rho)$

$$f = \frac{1}{2\pi\sigma^2 \sqrt{1-\rho}} \exp$$

$$\left[-\frac{1}{2(1-\rho^2)} \left\{\frac{(x-\mu_1)^2 + (y-\mu_2)^2 - 2\rho(x-\mu_1)(y-\mu_2)}{\sigma^2}\right\}\right]$$

$$L = \text{constant} - n \log \sigma^2 - \frac{n}{2} \log (1-\rho^2) - \frac{1}{2\sigma^2 (1-\rho^2)}$$

$$\left[\sum_i (x_i - \mu_1)^2 + \sum (y_i - \mu_2)^2 - 2\rho \sum (x_i - \mu_1)(y_i - \mu_2)\right]$$

$$\frac{\partial L}{\partial \mu_1} = \frac{1}{\sigma^2 (1-\rho^2)} \left[\sum_i (x_i - \mu_1) - \rho \sum_i (y_i - \mu_2)\right] = 0$$

$$\frac{\partial L}{\partial \mu_2} = \frac{1}{\sigma^2 (1-\rho^2)} \left[\sum_i (y_i - \mu_1) - \rho \sum_i (x_i - \mu_1)\right] = 0$$

yield $\quad (\bar{x} - \mu_1) = \rho (\bar{y} - \mu_2)$

$\quad (\bar{y} - \mu_2) = \rho (\bar{x} - \mu_1).$

Solving simultaneously $\hat{\mu}_1 = \bar{x},\ \hat{\mu}_2, = \bar{y}$ (1)

$$\frac{\partial L}{\partial \sigma^2} = -\frac{n}{\sigma^2} + \frac{1}{2\sigma^4 (1-\rho^2)} \cdot$$

$$\left[\sum_i (x_i - \mu_1)^2 + \sum_i (y_i - \mu_2)^2 - 2\rho \sum_i (x_i - \mu_1)(y_i - \mu_2)\right]$$

$$= 0$$

$$2n\sigma^2 (1-\rho^2) = \sum_i (x_i - \mu_1)^2 + \sum_i (y_i - \mu_2)^2 - 2\rho \sum_i (x_i - \mu_1)(y_i - \mu_2) \tag{2}$$

$$\frac{\partial L}{\partial \rho} = \frac{n\rho}{1-\rho^2} - \frac{\rho}{\sigma^2 (1-\rho^2)^2} \cdot$$

$$\left[\sum_i (x_i - \mu_i)^2 + \sum_i (y_i - \mu_2)^2 - 2\rho \sum_i (x_i - \mu_1)(y_i - \mu_2)\right]$$

$$+ \frac{\sum (x_i - \mu_1)(y_i - \mu_2)}{\sigma^2 (1-\rho^2)} = 0$$

$$n\rho (1-\rho^2) \sigma^2$$

$$= \rho \left[\sum_i (x_i - \mu_1)^2 + \sum_i (y_i - \mu_2)^2 - 2\rho \sum_i (x_i - \mu_1)(y_i - \mu_2)\right]$$

$$- (1-\rho^2) \sum_i (x_i - \mu_1)(y_i - \mu_2) . \tag{3}$$

Multiplying (2) by ρ and subtracting (3),

$$n\rho\,\sigma^2 = \sum_i (x_i - \mu_1)(y_1 - \mu_2) \tag{4}$$

Substituting (4) in (2)

$$2n\,\sigma^2(1-\rho^2) = \sum_i (x_i-\mu_1)^2 + \sum (y_i-\mu_2)^2 - 2n\rho^2\sigma^2$$

We obtain $$\hat{\sigma}^2 = \frac{\sum_i (x_i-\bar{x})^2 + \sum_i (y_i-\bar{y})^2}{2n} \tag{5}$$

Substituting (5) in (4)

$$\hat{\rho} = \frac{\sum_i (x_i-\bar{x})(y_i-\bar{y})}{n\hat{\sigma}^2}.$$

Thus, $$\hat{\mu}_1 = \bar{x},\ \hat{\mu}_2 = \bar{y},\ \hat{\rho} = \frac{\sum_i (x_i-\bar{x})(y_i-\bar{y})}{n\hat{\sigma}^2}$$

$$\hat{\sigma}^2 = \frac{\sum_i (x_i-\bar{x})^2 + \sum_i (y_i-\bar{y})^2}{2n}.$$

A.11

BVN $(\mu, \mu, \sigma^2, \sigma^2, \rho)$.

$$L = \text{Constant} - n\log\sigma^2 - \frac{n}{2}\log(1-\rho^2) - \frac{1}{2\sigma^2(1-\rho^2)}\cdot$$
$$\left[\sum_i (x_i-\mu)^2 + \sum_i (y_i-\mu)^2 - 2\rho\sum_i (x_i-\mu)(y_i-\mu)\right]$$

$$\frac{\partial L}{\partial \mu} = \frac{1}{\sigma^2(1-\rho^2)}\cdot$$
$$\left[\sum_i (x_i-\mu) + \sum_i (y_i-\mu) - \rho\cdot\right.$$
$$\left.\left\{\sum_i (x_i-\mu) + \sum_i (y_i-\mu)\right\}\right] = 0$$

yields $\qquad \bar{x}+\bar{y}-2\hat{\mu} = 0$

$$\hat{\mu} = \frac{\bar{x}+\bar{y}}{2}.$$

Similarly, solving the equations

$\dfrac{\partial L}{\partial \sigma^2} = 0$ and $\dfrac{\partial L}{\partial \rho} = 0$, simultaneously, we have

$$\hat{\sigma}^2 = \frac{\sum_i (x_i-\hat{\mu})^2 + \sum_i (y_i-\hat{\mu})^2}{2n},$$

and

$$\hat{\rho} = \frac{\sum_i (x_i-\hat{\mu})\,(y_i-\hat{\mu})}{n\hat{\sigma}^2}.$$

REFERENCES

Abramowitz, M. and Stegun, I. (1964). *Handbook of Mathematical Functions.* Natl. Bur. Stand. Applied Mathematical Series 55, U.S. Government Printing Office.

Aitchison, J. (1964). ''Bayesian tolerance region'', *J. Roy, Statist. Soc.,* **B, 26,** 161-175.

Aitchison, J. and Brown, J.A.C. (1957). *The Lognormal Distribution.* Cambridge University Press, London.

Aitchison, J. and Dunsmore, I.R. (1975). *Statistical Prediction Analysis.* Cambridge University Press, London.

Anderson, T.W. (1968). *An Introduction to Multivariate Statistical Analysis,* 342. Wiley, New York.

Bain, L.J. (1978), *Statistical Analysis of Reliability and Life Testing Model.* Marcel Dekker, New York.

Banerjee, A.K. and Bhattacharyya, G.K. (1977). ''A purchase incidence model with Inverse Gaussian interdependence times''. *J. Amer. Statist Assoc.* **71**, 823-829.

Basu, A.P. (1964). ''Estimation of reliability for some life distributions useful in life testing''. *Technometrics,* **6**, 215-219.

Bayes, Rev. T.R. (1763). ''An essay towards solving a problem in the doctrine of chances''. *Phil. Trans. Roy. Soc.,* **53**, 370-418. Reprinted in *Biometrika,* **45**, 296-315.

Berger, James, O. (1985). *Statistical Decision Theory: Foundations, Concepts and Methods.* Springer-Verlag, New York.

Biegel, John, E. (1974). ''Inspection errors and sampling plans''. *Amer. Inst. Ind. Eng. Trans.,* **6**, 284-287.

Box, G.E.P. and Tiao, G.C. (1973). *Bayesian Inference in Statistical Analysis,* Addison-Wesley, New York.

Cheng, R.C.H. and Amin, N.A.K. (1981). "Maximum likelihood estimation of parameters with Inverse Gaussian distribution with unknown origin". *Technometrics*, **23**, 257-263.

Chhikara, R.S. and Folks, J.L. (1977). "The Inverse Gaussain distribution as a life testing model". *Technometrics,* **19**, 461-468.

Chhikara, R.S. and Guttman, Irwin, (1982). "Predication limits for the Inverse Gaussian distribution". *Technometrics,* **24**, 319-324.

Cochran, W.G. (1963). *Sampling Techniques*, 2nd ed., Wiley, New York.

Cohen, A.C. and Whitten, B. (1982). "Modified maximum likelihood and modified moment estimators for the three-parameter Weibull distribution". *Comm. Statistist.*, **11 (23)**, 263-265.

De Finetti, B. (1937). "*La Prévision*: *Ses Lois Logiques Ses Sources Subjectives*"; translated into English by H.E. Kyberg, Jr. and H.G. Smokler (1964) in *Studies in Subjective Probability*. Wiley, New York.

DeBrujin, N.G. (1961). *Approximate Methods in Analysis*. North Hollander, Amsterdam.

DeGroot, Morris H. (1970). *Optimal Statistical Decisions,* McGraw-Hill, New York.

Dudewicz, E.J. (1976). Introduction to Statistics and Probability. Holt Reinhart.

Dumonceaux, R. and Antle, C.E. (1973). "Discrimination between the lognormal and Weibull distributions". *Technometrics,* **15**, 923-926.

Dunsmore, I.R. (1974). "The Bayesian predictive distribution in life testing models". *Technometrics,* **16**, 455-460.

Dyer, D.D. and Whisenand, C.W. (1973). "Best linear unbiased estimator of the parameter of Rayleigh distribution". *IEEE Trans.*, **R-22**, 229-231.

Edwards, W., Lindman, H., and Savage, L.J. (1963). "Bayesian Statistical Inference for Psychological research". *Psychological Rev.*, **70**, 193.

El-Sayyad, G.M. (1967). "Estimation of the parameters of an exponential distribution". *J. Roy. Statist. Soc.,* **B. 29**, 525-532.

Evans, I.G. (1976). Unpublished Lecture Notes. University College of Wales, Aberyswyth, U.K.

Evans, I.G. and Nigm, A.H.M. (1980). "Bayesian prediction for the left-truncated exponential distribution". *Technometrics,* **22**, 201-204.

Ferguson, Thomas S. (1967). *Mathematical Statistics: A Decision Theoretic Approach*. Academic Press, New York.

Fisher, Sir Ronald A. (1936). ''Uncertain inference''. *Proc. Amer. Acad. Sci.* **71**, 245.

Folks, J.L. and Chhikara, R.S. (1978). ''The Inverse-Gaussian distribution and its applications—a review''. *J. Roy. Statist. Soc.*, **B, 40**, 269-289.

Forgionne, G.A. (1986). *Quantitative Decision Making*. Wadsworth, Belmont, California.

Gren. J. Discussant on D.V. Lindley's (1980) paper ''Approximate Bayesian Method''. *Trabajos Estadistica,* **31**, 241.

Grubbs, F.E. (1971). ''Approximate fiducial bounds on reliability for two-parameter negative exponential distribution''. *Technometrics,* **13**, 873-876.

Guttman, Irwin (1970). *Statistical Tolerance Regions: Classical and Bayesian.* Griffin's Statistical Monographs. London, U.K.

Guttman, Irwin and Tiao, G.C. (1964). ''A Bayesian approach to some best population problems''. *Ann. Math. Statist.,* **35**, 825-835.

Guttman, Irwin, Wilks, S.S. and Hunter, J. Stuart (1971). *Introductory Engineering Statistics.* Wiley, New York.

Hapuarachchi, K.P. and Sinha, S.K. (1990). ''Sampling plans for bulk material''. *Sankhya,* Series **B**., Vol. **52**, 190-197.

Hartigan, J.A. (1964). ''Invariant prior distributions''. *Ann. Math. Statist.,* **35**, 836-845.

Herbach, L.H. (1959). ''Properties of Type II Analysis of variance tests''. *Ann. Math. Stat.*, **30**, 939-959.

Howlader, H.A. and Sinha, S.K. (1984). ''Bayesian estimation of regression parameters under a bivariate normal prior''. *J. Inst. Statist. Statistic Mathematique*, **29**, 47-57.

Jedamus, Paul, Frame, Robert and Taylor, Robert (1976). *Statistical Analysis for Business Decisions*, McGraw Hill, New York.

Jeffreys, Sir Harold (1946). ''An invariant form for the prior probability in estimation problems''. *Proc. Roy. Soc. London*; Sec. **A**, **186**, 453-461.

Jeffreys, Sir Harold (1961). *Theory of Probability,* 3rd. ed. Oxford, Clarendon Press.

Johnson, N.L. and Kotz, S. (1970). *Distributions in Statistics; Continuous Univariate Distributions I.* Houghton Muffin Series in Statistics. Boston.

Kendall, M.G. and Stuart, A. (1973). *The Advanced Theory of Statistics*, Vol. 2, 3rd ed. Hafner Publishing, New York.

Kenny, J.F. and Keeping, E.S. (1951), *Mathematics of Statistics*, Vol. 2. Van Nostrand, New York.

Keynes, J.M. (1921). *A treaties on probability*. Harper and Row, New York.

Kimeldorf, G.S. and Jones, D.A. (1967). "Bayesian graduation". *Trans. Soc. Act.* **19**, 66.

Kloek, T. and Van Dijk, H.K. (1978). "Bayesian estimation of equation system parameters: An application of integration by Monte Carlo". *Econometrika*, **46**, 1-19.

Lancaster, A. (1972). "A stochastic model for the duration of a strike". *J. Roy Statist. Assoc.*, **135**, 257-271.

Lawless, J.F. (1977). "Prediction intervals for the two-parameter exponential distribution". *Technometrics*, **19**, 469-472.

Lindley, D.V. (1965). *Introduction to Probability and Statistics from a Bayesian viewpoint*: Vol. 2, "Inference". Cambridge University Press, U.K.

Lindley, D.V. (1980). "Approximate Bayesian Method". *Trabajos Estadistica*, **31**, 223-237.

London, Dick (1985). *Graduation: The Revision of Estimates*, Actex Publication, Connecticut.

Lye, L., Sinha, S.K. and Booy, C. (1988). "Bayesian analysis for the *T*-year events for flood data fitted by a three-parameter lognormal distribution". *Civil. Eng. Sys.*, **5**, 81-86.

Mendenhall, W. and Hader, R.J. (1958). "Estimation of parameters of mixed exponentially distributed failure time distributions from censored failure data". *Biometrika*, **45**, 504-520.

Nadas, A. (1973). "Best tests for zero draft based on first passage times in Brownian motion". *Technometrics*, **15**, 125-132.

Naylor, J.C. and Smith, A.F.M. (1982). "Application of a metod for the efficient computation of posterior distributions". *Appld. Statist.*, **31**, 214-225.

Nelson, Wyne (1982). *Applied Life Data Analysis*. Wiley, New York.

Neter, John, Wasserman, William and Whitmore, G.A. (1973). *Fundamental Statistics for Business and Economics*, 4th ed. Allyn and Bacon, Boston.

Neden, Linda R. (1986). Bayesian one-side credibility bounds for a proportion using a two-sided sampling plan involving imperfect and perfect classfication. Unpublished M.Sc. Thesis, Department of Statistics, University of Manitoba, Winnipeg, Canada.

Padgett, W.J. (1981). ''Bayes estimation of reliability for Inverse Gaussian model''. *IEEE Trans. Rel.*, **R-30**, 384-385.

Polovko, A.M. (1968). *Fundamentals of Reliability Theory.* Academic Press, New York.

Pope, J., Lehrer, B. and Stephens, J.P. (1980). ''A multiphasing reading screening procedure''. *Jour. Learning Disabilities*, **13**, 98-102.

Raiffa, H. and Schlaifer, R. (1961). Applied Statistical Decision Theory. Harvard University Press.

Ramsey, F.P. (1931). *The foundation of Mathematics and other logical essays.* London: Routledge and Kegan Paul Ltd.

Rockette, H., Antle, C. and Klimko, L.A. (1974). ''Maximum likelihood estimation with the Weibull model''. *Jour Amer. Statist. Assoc.*, **69**, 246-249.

Samanta, M. and Sinha, S.K. (1979), ''Efficient estimation of reliability of two-component system''. *Trans. Ind. Assoc. for Prod. Quality and Reliability*, **4**, 77-84.

Savage, L.J. (1954). *The Foundation of Statistics.* Wiley, New York.

Savage, L.J. (1961). ''The subjective basis of statistical practice''. Technical Report, Department of Statistics, University of Michigan, Ann Arbor.

Savage, L.J. (1962). *The Foundation of Statistical Inference: A Discussion.* Methuen, London.

Siddiqui, M.M. (1962). ''Some problems connected with Rayleigh distribution''. *Jour. Res. Nat. Bur. Standard,* **661**, No. **2**, 167-174.

Sinclair, M.A. (1978). ''A collecltion of modelling approaches for visual inspection in industry''. *Int. Jour. Prod. Res.*, **16**, 275-292.

Sinha, S.K. (1980). ''Robustness of posteriors under improper priors in some life testing models''. *Jour. Ind. Statist. Assoc.*, **18**, 127-144.

Sinha, S.K. (1982), ''Maximum likelihood estimation of the reliability function with Type I censoring form a mixture of Weibull distributions''. *Jour. Inst. Statist: Statistique Mathematique*, **27**, 75-92.

Sinha, S.K. (1983). ''Bayesian estimation of the reliability function of mixed exponentially distributed time-censored life test data''. *Commun. Statist.*, **12**, 141-151.

Sinha, S.K. (1985). ''Bayes estimation of reliability function of normal life distribution''. *IEEE Trans. Rel.*, **R-34**, 60-74.

Sinha, S.K. (1986). *Reliability and Life Testing.* Wiley Eastern Ltd, Delhi, India.

Sinha, S.K. (1987). "Bayesian estimation of parameters and reliability function of a mixture of Weibull distributions". *Jour. Statist. Planning and Inf.*, **16**, 377-387.

Sinha, S.K. (1980). "On the prediction limits for Rayleigh distributions". *Cal. Statist. Bull.*, **39**, 105-109.

Sinha, S.K. (1995). "Prediction internal and reliability assessment of the kth failure time in exponential life distribution". To appear.

Sinha, S.K. and Guttman, Irwin (1988). "Bayesian analysis of life testing problems involving the Weibull distribution". *Commun. Statist.*, **A17**, 343-356.

Sinha, S.K. and Guttman, Irwin (1994). "Misclassification in Acceptance Sampling — A Bayesian Approach". *Jour. Indian Statistical Association*, **32**, No. 2, 83-94.

Sinha, S.K. and Howlader, H.A. (1984). "Bayesian estimation of regression parameters under a bivariate normal prior". *Jour. Inst. Statist: Statistique Mathematique*, **29**, 47-57.

Sinha, S.K. and Kale, B.K. (1980). *Life Testing and Reliability Estimation*. Wiley Eastern/Halsted Press, Delhi/New York.

Sinha, S.K. and Sloan, J.A. (1988). "Bayesian estimation of the parameters and reliability function of a three-parameter Weibull life distribution". *IEEE Trans. Rel.*, **R-27**, 364-369.

Sinha, S.K. and Sloan, J.A. (1989). "Prediction interval for a mixture of Weibull failure time distribution". *S. African J. of Statist.*, **23**, 119-130.

Sinha, S.K. and Zellner, Arnold (1990). "A note on the prior distributions of Weibull parameters". *Journal of the Society of Management Science and Applied Cybernatics (SCIMA)*, Vol. **19**, 5-13.

Sloan, J.A. and Sinha, S.K. (1990). "Bayesian predictive interval for a mixture of exponential failure time distribution with censored samples". *Statist. Prob. Letters.*, Vol. **11**, No. **6**, 537-545.

Sloan, J.A. and Sinha, S.K. (1995), "Bayesian estimation for bi-variate normal distribution" Bayesian analysis in statistics and economics–Essays in honour of Professor Arnold Zellner. John Wiley and sons, Chapter 10, 115-127.

Spurr, W.A. and Bonini, C.P. (1973). *Statistical Analysis for Business Decisions*. Richard Irwin, Homewood, Illinois.

Tanner, L. and Lerner, M. (1951). ''Economic accumulation of variance data in connection with bulk sampling''. *Amer. Soc. Testing and Material*, **114**, 8-12.

Tierney, L. and Kadane, J.B. (1986). ''Accurate approximations for posterior moments and marginal densities''. *J. Amer. Statist. Assoc.*, **81**, 82-86.

Tweedie, M.C.K. (1957). ''Statistical properties of Inverse Gaussian distribution-1''. *Ann. Malth. Stat.*, **28**, 362-377.

Uspensky, J.V. (1937). *Introduction to Mathematical Probability.* McGraw-Hill, New York.

Varian, H.R. (1975). ''A Bayesian approach to real estate assessment''. *Studies in Bayesian Econometrics and Statistics in Honor of Leonard J. Savage*, 195-208. (Eds.) Stephen E. Feinberg and Arnold Zellner. North Holland, Amsterdam.

Weibull, W. (1939). ''A statistical theory of the strength of material''. *Ing. Vetenskps. Akad. Handl.*, **151,** 1-45.

Whitmore, G.A. (1976). ''Management application of the Inverse Gaussian distribution''. *Int. J. Managt. Sc.*, **4**, 215-223.

Whitmore, G.A. (1986). ''Prediction limits for a univariate normal observation''. *Amer. Stat.*, **40,** 141-143.

Wilks, S.S. (1962). *Mathematical Statistics.* Wiley, New York.

Zellner, Arnold (1971). *An Introduction to Bayesian Inference in Econometrics*, Wiley, New York.

Zellner, Arnold (1986). ''Estimation and prediction using asymmetric loss functions''. *J. Amer. Statist. Assoc.*, **81**, 446-451.

Zellner, Arnold and Giesel, Martin S. (1968). ''Sensitivity of control to uncertainty and form of the criterion function''. *Future of Statistics,* (ed) Donald G. Watts. Academic Press, New York.

Zellner, Arnold and Ross, Peter E. (1984). ''Bayesian analysis of dichotomous quantal response model''. *J. of Econometrics*, **25**, 365-393. North-Holland, Amsterdam.

AUTHOR INDEX

SUBJECT INDEX

Of Related Interest

A LINEAR MODEL APPROACH TO REGRESSION ANALYSIS

SUDDHENDU BISWAS

The book develops the classical and modern unified approaches of linear models providing passage to Regression Analysis. While the theory of linear models have been treated exhaustively both by classical as well as by g-inverse approaches, the topic of Regression Analysis has been presented through examples. The write-up of the book has been selective to cover the course within the time constraint of a semester; unlike other exhaustive authoritative treatises on the subject which besides being less comprehensive do not meet the requirement of doing justice to the subject within the time constraint of a semester. Derivation of many results of least squares theory by the use of g-inverse and Projection matrices, development of the background of ANOVA and ANCOVA through linear estimation approach and learning process of Regression Analysis through a complete chapter of examples are some of the unique features of the book.

Salient Features : Fundamentals of Linear Estimation • Testing of Linear Hypothesis • Theory of Linear Estimation Based on g-inverse • Miscellaneous Problems in Least Squares Theory • Regression Theory and its Applications • Application of Linear Estimation Techniques

BIOMETRICAL AND STATISTICAL TECHNIQUES IN PLANT BREEDING

J.R. SHARMA

The book presents a comprehensive account of the concept and genesis of diverse biometrical/statistical models as applied to plant breeding experiments under different situations. Generation and statistical treatment of data; presentation, interpretation and inferences of results; merits, demerits and situations of applicability of models are all explicated for their adequate and appropriate usage in plant breeding. The whole volume comprising 25 chapters has been zipped into five sections elucidating; general statistical/biometrical parameters and field designs (Chapters 1-4), multivariate analysis of genetic divergence (Chapters 6-7), genotype x environment interaction and stability parameters (Chapters 8-10), analysis of nature of gene action and variance components (Chapters 11-23), and lastly the unique analysis of statistical and genetical parameters related to selection and mutation experiments (Chapters 24-25) in plant breeding.

Simplification of the bewildering complexities of biometrical notations and procedures in a language which could easily be grasped by biologists/geneticists having little or no statistical background is the hallmark of the treatise. Like a ready-reckoner, this work offers an efficient key to plant breeding data-management for both students and professional plant breeders alike in pursuit of their research goals.

PUBLISHING FOR ONE WORLD

NEW AGE INTERNATIONAL (P) LIMITED, PUBLISHERS

New Delhi • Bangalore • Calcutta • Chennai • Guwahati • Hyderabad
Lucknow • Mumbai • Pune

ISBN: 81-224-1171-1